Cuk-H03679-8-P024839

⑧

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

P, 24 839

EDITORIAL COMMITTEE (1978)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

E. W. HANCOCK
Stanford University

K. J. ISSELBACHER
Massachusetts General Hospital

S. L. KATZ
Duke University

P. G. QUIE
University of Minnesota

E. J. SACHAR
Albert Einstein College of Medicine

G. T. SHIRES
Cornell University

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

Responsible for the organization of Volume 29 (Editorial Committee, 1976)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

P. B. GREGORY (Guest)
Stanford University

E. W. HANCOCK
Stanford University

K. J. ISSELBACHER
Massachusetts General Hospital

R. M. ROSE (Guest)
Boston University

E. J. SACHAR
Albert Einstein College of Medicine

J. P. SANFORD
The University of Texas

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

R. T. SMITH
The University of Florida

L. W. WANNAMAKER
The University of Minnesota

Production Editors
JULIE FAGAN
MARGOT PLATT

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

WILLIAM P. CREGER, *Editor*
Stanford University School of Medicine

C. H. COGGINS, *Associate Editor*
Harvard Medical School

E. W. HANCOCK, *Associate Editor*
Stanford University School of Medicine

VOLUME 29

1978

ANNUAL REVIEWS INC. 4139 EL CAMINO WAY PALO ALTO, CALIFORNIA 94306

ANNUAL REVIEWS INC.
Palo Alto, California, USA

COPYRIGHT © 1978 BY ANNUAL REVIEWS INC., PALO ALTO, CALIFORNIA, USA. ALL RIGHTS RESERVED. The appearance of the code at the bottom of the first page of an article in this serial indicates the copyright owner's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per-copy fee of $1.00 per article through the Copyright Clearance Center, Inc. (P.O. Box 765, Schenectady, NY 12301) for copying beyond that permitted by Sections 107 or 108 of the US Copyright Law. The per-copy fee of $1.00 per article also applies to the copying, under the stated conditions, of articles published in any Annual Review serial before January 1, 1978. Individual readers, and nonprofit libraries acting for them, are permitted to make a single copy of an article without charge for use in research or teaching. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

REPRINTS The conspicuous number aligned in the margin with the title of each article in this volume is a key for use in ordering reprints. Available reprints are priced at the uniform rate of $1.00 each postpaid. The minimum acceptable reprint order is 10 reprints and/or $10.00 prepaid. A quantity discount is available.

International Standard Serial Number: 0066-4219
International Standard Book Number: 0-8234-0529-9
Library of Congress Catalog Card Number: A51-1659

Annual Reviews Inc. and the Editors of its publications assume no responsibility for the statements expressed by the contributors to this Review.

PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

Annual Review of Medicine
Volume 29, 1978

CONTENTS

ALLERGY AND IMMUNOLOGY:

Polymyalgia Rheumatica and Giant Cell Arteritis, *Robert E. Ettlinger, Gene G. Hunder, and L. Emmerson Ward* 15

ANESTHESIA:

Biology of Drowning, *Jerome H. Modell* 1

CARDIOVASCULAR:

Left Ventricular Stiffness, *Judith A. Wisneski and J. David Bristow* 475
Cardiogenic Hypertension, *Harriet P. Dustan and Robert C. Tarazi* 485
Cardiac Assist Devices, *John J. Lamberti and Leon Resnekov* 571
Slow-Channel Depolarization: Mechanism and Control of Arrhythmias, *John C. Bailey, Victor Elharrar, and Douglas P. Zipes* 417

CLINICAL PHARMACOLOGY:

Neural Actions of Digitalis, *Richard A. Gillis and John A. Quest* 73
The Disposition and Effects of Sedatives and Analgesics in Liver Disease, *Anastacio M. Hoyumpa, Jr., Robert A. Branch, and Steven Schenker* 205

ENDOCRINOLOGY AND METABOLISM:

Control of the Onset of Puberty, *Robert M. Boyar* 509
Clinical Hypophosphatemia, *Faith Fitzgerald* 177
Medical Treatment of Acromegaly, *Ira D. Goldfine* 407
Osteomalacia and Disorders of Vitamin D Metabolism, *Joel F. Habener and Jane E. Mahaffey* 327
Humoral Syndromes Associated with Cancer, *W. D. Odell and A. R. Wolfsen* 379
Prostaglandins and Hypercalcemic States, *Hannsjörg W. Seyberth, Lawrence G. Raisz, and John A. Oates* 23

GASTROINTESTINAL TRACT:

Cellular Mechanisms in Acid Secretion, *Andrew H. Soll and Morton I. Grossman* 495

Digestion and Absorption of Protein, *Hugh J. Freeman and Young S. Kim* 99
Gastrointestinal Hormones, *James E. McGuigan* 307
Non-A, Non-B Hepatitis, *Stephen M. Feinstone and Robert H. Purcell* 359

HEMATOLOGY:

Erythropoietin and the Control of Red Cell Production, *Stanley E. Graber and Sanford B. Krantz* 51
Relationships between Platelets and Coagulation Factors in Hemostasis, *Philip W. Majerus and Joseph P. Miletich* 41
Heparin, Antithrombin, and Abnormal Clotting, *Robert D. Rosenberg* 367
Contractile Proteins in Cell Structure and Function, *Thomas P. Stossel* 427

INFECTION:

Chronic Bacteremia and Nephritis, *Youngki Kim and Alfred F. Michael* 319
Rabies Vaccination, *Stanley A. Plotkin and Tadeusz Wiktor* 583
Otitis Media: The Spectrum of Middle Ear Inflammation, *G. Scott Geibink and Paul G. Quie* 285
E. Coli Meningitis; K_1 Antigen and Virulence, *Catherine M. Wilfert* 129

KIDNEYS AND URINARY SYSTEM:

Dialysis Encephalopathy Syndrome, *Allen C. Alfrey* 93
Renal Tubular Acidosis, *F. John Gennari and Jordan J. Cohen* 521
Recurrent Glomerulonephritis after Renal Transplantation, *Jean Hamburger, Jean Crosnier, and Laure-Hélène Noël* 67
Immunological Aspects of Renal Tubular and Interstitial Diseases, *Robert T. McCluskey and Robert B. Colvin* 191
Dialysis Versus Transplantation in the Treatment of End-Stage Renal Disease, *John P. Merrill* 343

MISCELLANEOUS:

Informed Consent, *George J. Annas* 9
Recent Membrane Research and Its Implications for Clinical Medicine, *Vincent T. Marchesi* 593

NEOPLASIA:

Chemotherapy in Prevention of the Recurrence of Resectable Cancer, *Gianni Bonadonna, Gabriele Tancini, Anna Rossi, and Marco Gasparini* 149
The Non-Hodgkin's Lymphomas: Current Concepts and Management, *Carol S. Portlock and Eli Glatstein* 81
Immunotherapy for Malignant Disease, *James E. Goodnight, Jr. and Donald L. Morton* 231

PSYCHIATRY:

Anorexia Nervosa: Recent Investigations, *Katherine A. Halmi* 137

Behavioral Risk Factors in Coronary Artery Disease, *C. David Jenkins* 543

Genetic Transmission of Schizophrenia, *Dennis K. Kinney and Steven Matthysse* 459

Specificity of Depressive Diseases, *Donald R. Sweeney and James W. Maas* 219

The XYY Genotype, *Stanley Walzer, Park S. Gerald, and Saleem A. Shah* 563

REPRODUCTION:

The Prenatal Diagnosis of Genetic Disorders, *Charles J. Epstein and Mitchell S. Golbus* 117

Cardiovascular Complications of Oral Contraceptives, *Norman M. Kaplan* 31

CUMULATIVE INDEX OF CONTRIBUTING AUTHORS, VOLUMES 25–29 605

CUMULATIVE INDEX OF CHAPTER TITLES, VOLUMES 25–29 608

ANNUAL REVIEWS INC. is a nonprofit corporation established to promote the advancement of the sciences. Beginning in 1932 with the *Annual Review of Biochemistry,* the Company has pursued as its principal function the publication of high quality, reasonably priced Annual Review volumes. The volumes are organized by Editors and Editorial Committees who invite qualified authors to contribute critical articles reviewing significant developments within each major discipline.

Annual Reviews Inc. is administered by a Board of Directors whose members serve without compensation.

BOARD OF DIRECTORS
1978

Dr. J. Murray Luck
Founder Emeritus, Annual Reviews Inc.
Department of Chemistry
Stanford University

Dr. Joshua Lederberg
President, Annual Reviews Inc.
Department of Genetics
Stanford University Medical School

Dr. James E. Howell
Vice President, Annual Reviews Inc.
Graduate School of Business
Stanford University

Dr. William O. Baker
President
Bell Telephone Laboratories

Dr. Sidney D. Drell
Deputy Director
Stanford Linear Accelerator Center

Dr. Eugene Garfield
President
Institute for Scientific Information

Dr. William D. McElroy
Chancellor
University of California, San Diego

Dr. William F. Miller
Vice President and Provost
Stanford University

Dr. John Pappenheimer
Department of Physiology
Harvard Medical School

Dr. Colin S. Pittendrigh
Director
Hopkins Marine Station

Dr. Esmond E. Snell
Department of Microbiology
University of Texas at Austin

Dr. Harriet Zuckerman
Department of Sociology
Columbia University

Annual Reviews are published in the following sciences: Anthropology, Astronomy and Astrophysics, Biochemistry, Biophysics and Bioengineering, Earth and Planetary Sciences, Ecology and Systematics, Energy, Entomology, Fluid Mechanics, Genetics, Materials Science, Medicine, Microbiology, Neuroscience, Nuclear Science, Pharmacology and Toxicology, Physical Chemistry, Physiology, Phytopathology, Plant Physiology, Psychology, and Sociology. In addition, two special volumes have been published by Annual Reviews Inc.: *History of Entomology* (1973) and *The Excitement and Fascination of Science* (1965).

ERRATA

Volume 28 (1977)
In Treatment of Idiopathic Thrombocytopenic Purpura (ITP) by Yeon S. Ahn and William J. Harrington:
page 301, item (*d*) in the third paragraph should read autoimmune;
page 302, the name in the third line should be Heinz.

Ann. Rev. Med. 1978. 29:1–8
Copyright © 1978 by Annual Reviews Inc. All rights reserved

BIOLOGY OF DROWNING

◆7259

Jerome H. Modell, M.D.
Department of Anesthesiology, University of Florida College of Medicine,
Gainesville, Florida 32610

INTRODUCTION

Drowning has accounted for over 7000 fatalities annually in the past 25 years. A recent report showed that 89% of those who nearly drowned and were brought to the hospital survived with therapy (1). Extrapolating these figures, it can be estimated that more than 1.2 million patients were treated for near-drowning in the past 25 years. A clear understanding of the biologic changes that occur after near-drowning is important for successful therapy. Although the early literature emphasized changes in serum electrolyte concentrations, blood volume, and cardiovascular function, more recent experiments and clinical observations confirm that reversing the pulmonary insult is of prime importance.

PATHOPHYSIOLOGY

Changes in Blood Gas, Acid-Base Balance, and Pulmonary Status

The single most important consequence of near-drowning is hypoxemia (2). Mammals respond to total immersion in liquid by holding their breath or closing their vocal cords (3). Modell and colleagues (4) have shown that 80% of dogs anesthetized with barbiturates could be resuscitated by mechanically ventilating their lungs with air five minutes after the onset of tracheal obstruction, even though their mean arterial oxygen tension (P_{aO_2}) was only 4 torr when resuscitative efforts were started.

In treating 91 human victims of near-drowning, this author and his colleagues have seen 10 patients (12%) who probably near-drowned without aspirating water (1). In these cases, mouth-to-mouth resuscitation was given promptly at the scene of the accident, and spontaneous ventilation returned rapidly. If victims of near-drowning without aspiration are ventilated artificially before circulation ceases and before irreversible damage to the central nervous system occurs, recovery will be dramatic and complete. If spontaneous ventilation begins while the patient is still under water, however, aspiration will occur and treatment will be more complicated.

0066-4219/78/0401-0001$01.00

Craig (5) has demonstrated the importance of hypoxia in underwater swimming. Studying the breaking point of breath-holding in human volunteers during simulated underwater swimming, he found that when exercise followed hyperventilation, the breaking point dropped to 85 seconds, and the alveolar carbon dioxide tension ($P_{A_{CO_2}}$) remained fairly constant at 49 torr. However, the alveolar oxygen tension ($P_{A_{O_2}}$) dropped to 45 torr. Craig postulated that if the swimmer hyperventilated before swimming underwater, the arterial carbon dioxide tension ($P_{a_{CO_2}}$) would remain low. The swimmer, therefore, could consciously suppress the urge to breathe. He then would develop cerebral hypoxia, lose consciousness, and begin to breathe while submerged.

If aspiration occurs in the near-drowning victim, even though the initial hypercarbia may be reversed rapidly by increasing the patient's minute alveolar ventilation, he is frequently left with persistent arterial hypoxemia and metabolic acidosis (6). A $P_{a_{O_2}}$ as low as 21 torr and a pHa as low as 6.77 have been reported in humans (1). The metabolic acidosis frequently requires administration of sodium bicarbonate; the arterial hypoxemia can be profound and will persist as long as significant intrapulmonary pathologic conditions continue.

Although hypoxia occurs after aspiration of either freshwater or seawater, the factors contributing to the hypoxia may be different. Since seawater is hypertonic, fluid is drawn from the plasma into the alveoli, causing fluid-filled but perfused alveoli (7) and a large intrapulmonary shunt. When experimental animals aspirate freshwater, although the fluid is absorbed rapidly into the circulation (8), the surface-tension properties of pulmonary surfactant are altered (9). As a result, the alveoli become unstable and atelectasis occurs, producing intrapulmonary shunting and hypoxemia (10). Other factors that might contribute to pulmonary edema and alteration of ventilation-to-perfusion ratio in these patients are pulmonary hypertension and cerebral hypoxia (11).

If the patient survives for at least 12 hours after near-drowning only to die later, the lungs frequently show evidence of bronchopneumonia, multiple abscesses, mechanical injury, and deposition of hyaline material in the alveoli (12).

Recent studies report that placing positive end-expiratory pressure (PEEP) on the airway of animals near-drowned with seawater will markedly increase the $P_{a_{O_2}}$, whether the animals breathe spontaneously or are ventilated mechanically (13). This suggests that PEEP increases functional residual capacity, thereby minimizing intrapulmonary shunt and ventilation-to-perfusion abnormalities, and promoting better oxygenation.

Similar studies conducted after freshwater aspiration demonstrated an improvement in $P_{a_{O_2}}$ when mechanical ventilation was combined with PEEP. However, no improvement in $P_{a_{O_2}}$ was seen when 10-cm H_2O PEEP was applied in spontaneously breathing animals (14). The difference in results of these two studies may be that in order to overcome the surfactant changes that occur after aspiration of freshwater, the alveoli must be inflated mechanically before PEEP can effectively maintain the alveoli in an inflated state. Since pulmonary surfactant is normal after seawater aspiration, the spontaneous respiratory movement of the animal might be sufficient to inflate the alveoli, and PEEP might help maintain them in a state of partial inflation. It has been demonstrated in human victims of near-drowning that titration of PEEP improves ventilation-to-perfusion ratios and arterial oxygenation, not only when the patient's ventilation is controlled, but also when he breathes spontaneously and receives intermittent mandatory breaths from a ventilator (13).

Blood Volume and Serum Electrolytes

Aspiration of hypo- or hypertonic fluid can also cause changes in other systems, the extent and direction of which depend upon the quantity and nature of the fluid aspirated. If 11 ml/kg or less of fluid is aspirated, it is unlikely that persistent changes other than those caused by fluid in the lungs will occur in any system (15). However, if larger quantities of a hypotonic fluid are aspirated, the blood volume increases in direct proportion to the quantity of fluid aspirated (8). Conversely, when a hypertonic fluid such as seawater is aspirated, blood volume decreases linearly as the quantity of fluid aspirated increases (7).

Most patients do not aspirate enough fluid to produce changes in blood volume great enough to threaten life. Thus, while it is important to note the effective circulating blood volume by recording the patient's central venous pressure and pulmonary capillary wedge pressure as appropriate, blood-volume changes requiring urgent treatment are not usually seen. When a significant decrease in effective circulating blood volume is evident, it is usually the result of loss of fluid from the vascular space into the lung.

Changes that occur in serum electrolyte concentrations after drowning and near-drowning vary with the type and volume of water aspirated and are inversely proportional to changes in blood volume. Experiments have shown that if dogs aspirate 22 ml/kg or less of freshwater, significant persistent changes in extracellular serum electrolytes do not occur (8). Another study has demonstrated that less than 15% of human drowning victims who died in the water had serum electrolyte concentrations indicating aspiration of more than 22 ml/kg of water (15).

This author and his colleagues studied electrolyte concentrations in 83 patients who near-drowned in either freshwater, seawater, or brackish water, and in no case were the values found to be life-threatening (1). The ranges of values in the total group, regardless of the type of water aspirated, were as follows: serum sodium, 126–160 mEq per l; serum chloride, 86–126 mEq per l; and serum potassium, 2.4–6.3 mEq per l. This suggests that the importance of immediate therapy for electrolyte changes after near-drowning has been overstated in the past, and that each patient's electrolyte status should be evaluated before corrective therapy is initiated.

Hemoglobin and Hematocrit Values

When the hemoglobin and hematocrit values of these 83 near-drowning victims were analyzed, the mean value for hemoglobin was 13.1 g/100 ml and 13.4 g/100 ml for those near-drowned in freshwater and seawater, respectively (1). Likewise, there was no difference in mean hematocrit concentration between the two groups of patients.

Cardiovascular System

Most human near-drowning victims show remarkable stability of the cardiovascular system after resuscitation. Experimental evidence suggests that changes in cardiovascular function during near-drowning are caused predominantly by changes in $P_{a_{O_2}}$ and acid-base balance (1, 6–8, 10). Variations in blood volume and serum electrolyte concentrations can also contribute to cardiovascular changes. However, less than 15% of drowning victims aspirate sufficient water to cause such changes (15).

Neurologic and Renal Functions

The fear of resuscitating a near-drowning victim who may have a prolonged neurologic deficit is always present. However, there are reports of patients being resuscitated after submersion in freshwater for 10 (16), 20 (16), 22 (17), and 40 (18) minutes; and in seawater for 17 minutes (19). All of these patients were hypothermic; three of them developed prolonged neurologic deficits. Approximately one year later, they all functioned normally.

The cases just cited indicate the tremendous variation that exists in the individual's ability to tolerate submersion and subsequent hypoxemia, and should encourage rescuers to maintain resuscitative efforts as long as possible.

Adequate renal function remains intact in most patients who are resuscitated after near-drowning, although albuminuria, hemoglobinuria, oliguria, or anuria progressing to acute tubular necrosis can occur (12, 17, 19–22).

THERAPY

The primary objective of emergency therapy for the near-drowning victim is to restore arterial blood gas and acid-base levels to normal as rapidly as possible. Thus, immediate rescue and initial management of the near-drowned victim are crucial in determining the outcome. Since the degree of hypoxia increases rapidly with each second of apnea, it is imperative that emergency measures be taken immediately to initiate ventilation. There is little question that the most effective methods available today are still mouth-to-mouth and mouth-to-nose ventilation (23).

If the victim has not aspirated water, and if effective ventilation is established before permanent circulatory and neurologic changes occur, the prognosis is excellent. If water has been aspirated, persistent alterations of pulmonary function will probably occur. Time should not be wasted trying to drain water from the lungs of a victim of near-drowning in freshwater (24). Although it has been shown that the survival rate of animals aspirating large quantities of seawater can be increased if their lungs are drained by gravity (13), human drowning victims rarely aspirate such quantities. Therefore, it is more important to initiate artificial ventilation promptly than to drain the fluid.

On the other hand, victims of drowning and near-drowning frequently do swallow large quantities of water before losing consciousness. Therefore, it is important to ensure a clear airway, so that the stomach is not overdistended during resuscitation. If distention does occur, the patient may regurgitate and aspirate acid gastric contents.

When the proper equipment for mechanical ventilation is available, it should replace mouth-to-mouth resuscitation. The appliance or mechanical ventilator selected should be capable of producing PEEP.

If the patient does not have an effective heartbeat, closed-chest cardiac massage must be instituted immediately and an EKG obtained to determined whether electrical defibrillation is indicated.

Even if the patient breathes spontaneously after rescue or after receiving initial resuscitative efforts, the rescuer should not be lulled into a false sense of security. A patient who is able to converse with the rescuer may still have an extremely low P_{aO_2}. Supplemental oxygen therapy should be continued until actual measurement of P_{aO_2} confirms it to be unnecessary.

All near-drowning victims should be taken to a hospital for further evaluation and therapy. During transport, ventilatory and circulatory assistance should be provided as indicated.

Initial hospital therapy should emphasize pulmonary care, which may range from simply increasing the fractional concentration of inspired oxygen (F_{IO_2}) in a spontaneously breathing patient to providing continuous ventilatory support by establishing a patent airway with an endotracheal tube that is connected to a mechanical ventilator. At the present time, strict guidelines that establish conditions for which endotracheal intubation is necessary are not available. It has been shown in a series of 20 near-drowning victims, all of whom had normal chest roentgenograms on admission to the hospital, that only one required endotracheal intubation, even though some patients had P_{aO_2}'s as low as 52 torr (1).

Approximately 70% of near-drowning victims have a significant degree of metabolic acidosis (1). Therefore, we recommend that if the patient is unresponsive and if arterial blood-gas results are not immediately available, sodium bicarbonate (1.0 mEq/kg) should be given empirically. Of course, samples of arterial blood must be taken as soon as possible to evaluate oxygen and carbon dioxide tensions, pH, and bicarbonate level. These values determine the pattern of ventilatory support, the appropriate F_{IO_2}, and the amount of additional bicarbonate to be administered.

The exact amount of PEEP that is applied to any patient's airway should be determined on an individual basis, the level of PEEP being increased gradually. Its effect on oxygen tension and cardiovascular function should be assessed. The optimal level will be reached when PEEP is producing the lowest intrapulmonary venous admixture without adversely affecting cardiac output.

Many clinicians fear that increasing PEEP always decreases cardiac output. This can occur in the presence of hypovolemia. However, if a normal effective circulating blood volume is maintained, cardiac output may actually increase as PEEP increases and as P_{aO_2} improves (25).

If bronchospasm is present because of aspiration of fluid, it can be treated by administering an aerosol of bronchodilating agent, such as racemic epinephrine. The pulmonary edema frequently seen after near-drowning with aspiration can usually be controlled by applying PEEP.

Until recently, many physicians advocated the use of steroids and prophylactic broad-spectrum antibiotics to treat near-drowning victims (6, 26). Evidence now suggests that neither steroid nor prophylactic antibiotic therapy improves arterial oxygenation or survival rate in near-drowning victims (1). A better approach is to culture secretions from the trachea on a daily basis and to treat infection with specific antibiotics when it occurs, rather than to administer them prophylactically.

Many near-drowning victims regurgitate and may aspirate solid debris or particles of undigested food. Serial physical examinations and analyses of chest roentgenograms will help in diagnosing regional or lobar atelectasis caused by aspiration of particulate matter. If regional atelectasis occurs, bronchoscopic examination is indicated.

The patient's hematocrit value should be determined early in the therapy. Unless this value is markedly abnormal, or unless obvious hemolysis is present in the plasma fraction after freshwater aspiration, the problems in treating the patient will be limited almost exclusively to ventilation, oxygenation, and acid-base balance; significant fluid and electrolyte disturbances are unlikely to occur.

All near-drowning victims should be monitored closely. At the minimum, every patient should have vital signs, i.e. pulse, respiration, blood pressure, and temperature, monitored, and should have serial determinations made of arterial blood-gas tensions and pH. Electrocardiographic monitoring and urine output should be observed closely in patients who require any type of prolonged support. Arterial catheters facilitate blood sampling for reliable determination of pH and blood-gas tensions. In addition, if the patient shows any degree of circulatory instability, venous pressure should be monitored. Central venous pressure is only an approximate guide in the complete assessment of cardiac function and blood-volume balance, since it reflects only what may be happening proximal to the right atrium. It is frequently necessary to separate the effects of hypovolemia from those of cardiac failure when a low cardiac output and hypotension are present. These conditions require placing a pulmonary artery catheter so that pulmonary artery and pulmonary capillary wedge pressures can be obtained (27). The latter value is a better indicator of the function of the left side of the heart and is therefore a more useful guide in determining whether the patient requires additional fluid or supplemental cardiac support. Such a catheter also permits analysis of mixed venous blood, so that the arterial-venous oxygen-content difference can be monitored.

In addition, laboratory evaluation should include hemoglobin, hematocrit, and serum electrolyte determinations; urinalysis; culture of tracheal secretions; and chest roentgenograms. These evaluations should be made routinely, and other tests added as appropriate. Any abnormalities that are observed should then be corrected.

Deliberately induced hypothermia (16) has been advocated occasionally in the treatment of the drowning victim. The rationale is that hypothermia would decrease cerebral oxygen consumption. However, to be effective, hypothermia should be induced before hypoxia occurs. Obviously, this is not always possible in the near-drowning victim. As a result, the rationale for inducing hypothermia is questionable.

Some physicians have also recommended exchange transfusions because of a potentially high plasma hemoglobin level that might affect the kidneys adversely (17). Since patients with levels in excess of 500 mg/100 ml have not been reported (6), exchange transfusion is probably not necessary and only further delays the institution of proper therapy.

CONCLUSION

In summary, the near-drowning victim must be treated immediately for ventilatory insufficiency, hypoxia, and the resulting acidosis. The cause and pathophysiologic changes of pulmonary insufficiency vary, depending on the type and volume of fluid aspirated. Success or failure of the overall resuscitative effort frequently depends on the adequacy of prompt emergency resuscitation and on effective intensive pulmonary care. Each patient should be evaluated and treated individually, since abnormalities of multiple organ systems can occur, their degree and form varying considerably from patient to patient (28).

Literature Cited

1. Modell, J. H., Graves, S. A., Ketover, A. 1976. Clinical course of 91 consecutive near-drowning victims. *Chest* 70:231–38
2. Modell, J. H., Gaub, M., Moya, F., Vestal, B., Swarz, H. 1966. Physiologic effects of near drowning with chlorinated fresh water, distilled water and isotonic saline. *Anesthesiology* 27:33–41
3. Lougheed, D. W., Janes, J. M., Hall, G. E. 1939. Physiological studies in experimental asphyxia and drowning. *Can. Med. Assoc. J.* 40:423–28
4. Modell, J. H., Kuck, E. J., Ruiz, B. C., Heinitsh, H. 1972. Effect of intravenous vs. aspirated distilled water on serum electrolytes and blood gas tensions. *J. Appl. Physiol.* 32:579–84
5. Craig, A. B. Jr. 1961. Causes of loss of consciousness during underwater swimming. *J. Appl. Physiol.* 16:583–86
6. Modell, J. H., Davis, J. H., Giammona, S. T., Moya, F., Mann, J. B. 1968. Blood gas and electrolyte changes in human near-drowning victims. *J. Am. Med. Assoc.* 203:337–43
7. Modell, J. H., Moya, F., Newby, E. J., Ruiz, B. C., Showers, A. V. 1967. The effects of fluid volume in seawater drowning. *Ann. Intern. Med.* 67:68–80
8. Modell, J. H., Moya, F. 1966. Effects of volume of aspirated fluid

during chlorinated fresh water drowning. *Anesthesiology* 27:662–72
9. Giammona, S. T., Modell, J. H. 1967. Drowning by total immersion. Effects on pulmonary surfactant of distilled water, isotonic saline, and sea water. *Am. J. Dis. Child.* 114:612–16
10. Modell, J. H., Moya, F., Williams, H. D., Weibley, T. C. 1968. Changes in blood gases and A-aDO₂ during near-drowning. *Anesthesiology* 29:456–65
11. Moss, G., Staunton, C., Stein, A. A. 1972. Cerebral etiology of the "shock lung syndrome." *J. Trauma* 12:885–90
12. Fuller, R. H. 1963. The clinical pathology of human near-drowning. *Proc. R. Soc. Med.* 56:33–38
13. Modell, J. H., Calderwood, H. W., Ruiz, B. C., Downs, J. B., Chapman, R. Jr. 1974. Effects of ventilatory patterns on arterial oxygenation after near-drowning in sea water. *Anesthesiology* 40:376–84
14. Ruiz, B. C., Calderwood, H. W., Modell, J. H., Brogdon, J. E. 1973. Effect of ventilatory patterns on arterial oxygenation after near-drowning with fresh water: A comparative study in dogs. *Anesth. Analg.* 52:570–76
15. Modell, J. H., Davis, J. H. 1969. Electrolyte changes in human drowning victims. *Anesthesiology* 30:414–20
16. Ohlsson, K., Beckman, M. 1964. Drowning—reflections based on two cases. *Acta Chir. Scand.* 128:327–39
17. Kvittingen, T. D., Naess, A. 1963. Recovery from drowning in fresh water. *Br. Med. J.* 5341:1315–17
18. Siebke, H., Breivik, H., Rød, T., Lind, B. 1975. Survival after 40 minutes' submersion without sequelae. *Lancet* 1:1275–77
19. King, R. B., Webster, I. W. 1964. A case of recovery from drowning and prolonged anoxia. *Med. J. Aust.* 1:919–20
20. Rath, C. E. 1953. Drowning hemoglobinuria. *Blood* 8:1099–1104
21. Munroe, W. D. 1964. Hemoglobinuria from near-drowning. *J. Pediatr.* 64:57–62
22. Redding, J. S., Pearson, J. W. 1964. Management of drowning victims. *GP* 29:100–4
23. Safar, P., Escarraga, L. A., Elam, J. O. 1958. A comparison of the mouth-to-mouth and mouth-to-airway methods of artificial respiration with chest-pressure arm-lift methods. *N. Engl. J. Med.* 258:671–77
24. Ruben, A., Ruben, H. 1962. Artificial respiration. Flow of water from the lung and the stomach. *Lancet* 1:780–81
25. Downs, J. B., Klein, E. F. Jr., Modell, J. H. 1973. The effect of incremental PEEP on P_{aO_2} in patients with respiratory failure. *Anesth. Analg.* 52:210–14
26. Sladen, A., Zauder, H. L. 1971. Methylprednisolone therapy for pulmonary edema following near drowning. *J. Am. Med. Assoc.* 215:1793–95
27. Swan, H. J. C., Ganz, W., Forrester, J., Marcus, H., Diamond, G., Chonette, D. 1970. Catheterization of the heart in man with use of a flow-directed balloon-tipped catheter. *N. Engl. J. Med.* 283:447–51
28. Modell, J. H. 1977. Drowning and near-drowning. In *Pulmonary Disorders in Children,* ed. E. Kendig, V. Chernick. pp.498–510. Philadelphia: Saunders. 3rd ed.

Ann. Rev. Med. 1978. 29:9–14
Copyright © 1978 by Annual Reviews Inc. All rights reserved

INFORMED CONSENT ◆7260

George J. Annas, J.D., M.P.H.
Boston University, Schools of Law and Medicine, Boston, Massachusetts 02215

There is probably more misinformation and misunderstanding in the medical community concerning the doctrine of informed consent than any other legal construct. This brief article outlines the theoretical basis of the doctrine and reviews recent developments in the law.

THE DOCTRINE

The doctrine itself is relatively simple, requiring physicians to make certain disclosures in lay language to their patients before subjecting them to risky procedures. These are:

1. a description of the proposed treatment;
2. alternatives to the proposed treatment;
3. inherent risks of death or serious bodily injury in the proposed treatment;
4. problems of recuperation that are anticipated; and
5. any additional information other physicians would disclose in similar circumstances.[1]

As with any legal rule, there are exceptions. Thus the physician need not generally disclose the above listed information:

1. in an emergency;
2. if the patient does not want to be informed;
3. if the procedure is simple, and the danger remote and commonly appreciated as remote; and
4. if in the physician's judgment it is not in the patient's best interest to know, e.g. when the information would so seriously upset the patient that he could not rationally make a decision.[1,2]

Functions of the Doctrine

Of the approximately 200 appellate decisions to date dealing with the issue of informed consent, fewer than 25% (45 cases) mention the basis on which the

[1] Cobbs v. Grant, 8 Cal. 3d 229, 502 P.2d 1 (1972).
[2] Canterbury v. Spence, 464 F.2d 772, 789 (D.C. Cir. 1972).

0066-4219/78/0401-0009$01.00

court found or failed to find informed consent necessary. Of those that do give a basis, 25 rely on the patient's right of self-determination. The other cases tend to rely on the fiduciary qualities of the doctor-patient relationship—17 cases base their conclusions on it, 4 cases use it in conjunction with a self-determination rationale.[3] Just as rationales vary from jurisdiction to jurisdiction, so do specific rules concerning the law of informed consent. Nevertheless, it seems correct to conclude that there are two primary functions of the doctrine of informed consent: to promote individual autonomy, and to encourage rational decision making.

The purpose of autonomy (the "right to be left alone,"[2] or "the right to privacy") is to protect the individual's personal integrity by denying anyone the right to invade his body without his consent.[3] Encouraging rational decision making is clearly viewed by courts as a secondary function of the informed consent doctrine. But, in fact, it is an extremely important function since, without rational decision making, the entire doctrine is called into serious question.

Hopefully, this self-scrutiny will prevent some procedures from being performed. For example, after a candid review of risks, some patients, based on their personal assessment, might refuse treatment. Informed consent in this instance helps insure that those who bear the bodily and psychological risks of a treatment will have the final decision as to whether or not it is performed.

The traditional arguments against the patient having a major role in the decision-making process are that 1. the patient will never be able to comprehend the information related; and 2. the information will unduly frighten the patient, and he will therefore not consent to a procedure that actually entails only a minimal risk.

It is the physician's duty to sufficiently inform and educate the patient to enable him to make up his own mind. If the physician argues this is not possible, he may in fact be saying one of two things: he cannot properly explain the risks and alternatives because he does not understand them himself, or he believes if he does properly explain them, the patient will not consent. In either case rational decision making will be promoted by the development of an adequate disclosure statement, and by the requirement that the patient be given final authority.

STATE LAWS ON INFORMED CONSENT

Prior to 1975 there were almost no state laws dealing specifically with this subject. In 1975 and 1976, however, 18 states,[4] in reaction to the malpractice insurance crisis,[3,4] enacted statutes that either defined or restricted the appli-

[3] Annas, G. J., Glantz, L. H. and Katz, B. K. 1977. *Informed Consent to Human Experimentation: The Subject's Dilemma,* p. 33. Cambridge, Mass.: Ballinger. 333 pp.

[4] Alaska, Colorado, Delaware, Florida, Idaho, Iowa, Kentucky, Nebraska, Nevada, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, Tennessee, Utah, Vermont, and Washington.

cation of the doctrine of informed consent. Most of the statutes attempt to make it more difficult for a patient to successfully sue a physician who failed to obtain informed consent.

Statutes in nine states, for example, provide that a patient's signature on a consent form shall be conclusive evidence that the information was provided to the patient and that the consent was valid.[5] Florida's statute provides that the signed consent shall be "conclusively presumed to be valid consent."[6] This presumption "may be rebutted if there was a *fraudulent misrepresentation of a material fact* in obtaining the signature"[6] (emphasis supplied in this and the three quotes that follow). Idaho's statute likewise provides that "such written consent, in the absence of convincing proof that it was *secured maliciously or by fraud,* is presumed to be valid . . . and the advice and disclosures of the attending physician or dentist, as well as the *level of informed awareness* of the giver of such consent, shall be presumed sufficient."[7] In Iowa:

> A consent in writing to any medical or surgical procedure or course of procedures *in patient care* which meets the requirements of this section *shall create a presumption* that informed consent was given.[8]

And in Ohio:

> Written consent to a surgical or medical procedure or course of procedures shall, to the extent that it fulfills . . . [certain requirements] *be presumed to be valid and effective,* in the absence of proof by a preponderance of the evidence that the person who sought such consent was *not acting in good faith,* or that the execution of the consent was induced by *fraudulent misrepresentation of material facts,* or that the person executing the consent was *not able to communicate effectively* in spoken or written *English* or any other language in which the consent is written.[9]

In Iowa and Ohio, statutes also contain lists of the types of risks that require disclosure by physicians. The language of each is set forth because it indicates the types of risks a legislature (as opposed to a court) finds to be significant to patients in the therapeutic setting. The lists are identical; both statutes require disclosure of the following risks: "death, brain damage, quadriplegia, paraplegia, the loss of function of any organ or limb, or disfiguring scars . . . with the probability of each such risk if reasonably determinable."[8,9] Statutes in eight other states are similar to New York's, in which the physician is required to disclose only those "risks and benefits involved as a reasonable medical practitioner under similar circumstances would have disclosed. . . ."[10] Five other states—Colorado, Utah, Nevada, Alaska, and Idaho—(in addition to Ohio and Iowa) adopt their own definitions of informed consent, while Pennsylvania and Washington base disclosure on the "risks and alternatives to treatment or diagnosis that a reasonable patient would consider material to the decision whether or not to undergo treatment or diagnosis."[3]

---

[5] Colorado, Florida, Iowa, Idaho, Nevada, North Carolina, Ohio, Utah, and Washington.

[6] Fla. Stats. Ann., sec. 768.132(4)(a).

[7] Idaho Code 439-4305.

[8] Iowa Code Ann. ch. 147 (added by H. B. 803, sec. 16, 1975).

[9] Ohio Rev, Code Ann. sec. 2317.54.

[10] Delaware, Florida, Nebraska, New York, North Carolina, Kentucky, Vermont, and Tennessee.

Two other provisions are of interest. In cases involving a risk of death or serious injury, Colorado attempts to clarify the type of disclosure required in terms of probability. The physician is required to quantify such risk by approximating "the percentage within two percent of the risk associated with the procedure, which percentage is either contained in a recognized medical publication or is measured by the experience of the physician out of a substantial number of the same or similar procedures."[11] The unstated, but probably intentional, implication is that risks having less than a 2% probability need not be disclosed. Ohio goes one step further, and is the only state legislature which sets forth a model consent form in its statute, providing that if this form is properly used, consent shall be presumed to be valid. Because of its uniqueness, the entire text of the form is set forth here:

> CONSENT FOR MEDICAL PROCEDURE AND ACKNOWLEDGMENT OF RECEIPT OF RISK INFORMATION
>
> State law requires us to obtain your consent to your contemplated surgery or other medical procedure. What you are being asked to sign is simply a confirmation that we have discussed your contemplated operation or medical procedure and that we have given you sufficient information upon which to make a decision whether to have the operation or medical procedure and any choice as to the type of operation or medical procedure of your own free will. We have already discussed with you the common problems or undesired results that sometimes occur. We wish to inform you, not to alarm you. If you wish, however, we can go into more elaborate details or more unlikely problems. If you do not, that is also your privilege. Please read the form carefully and check the appropriate boxes. Ask about anything that you do not understand. We will be pleased to explain it. I hereby authorize and direct ________________, with associate or assistants of his choice to perform the following surgical, diagnostic, or medical procedure on ________________, my ________________, as we have agreed upon.
>
> relationship
>
> ---
>
> I further authorize the doctors to perform any other procedure that in their judgment is advisable for my well being. Details of this operation have been explained to me. Alternative methods of treatment, if any, have also been explained to me as have the advantages and disadvantages of each. I am advised that though good results are expected, the possibility and nature of complications cannot be accurately anticipated and that therefore there can be no guarantee as expressed or implied either as to the result of surgery or as to cure.
>
> Degree and kind of risks known to be associated with this procedure, including anesthesia; each marked box indicates some risks that are associated with this procedure:

---

[11] Col. St. sec. 13-20-302(2).

Comments

☐ Death ____________
☐ Brain Damage ____________
☐ Quadriplegia (paralysis of all arms and legs) ____________
☐ Paraplegia (paralysis of both legs) ____________
☐ Loss of organ ____________
☐ Loss of an arm or leg ____________
☐ Loss of function of organ ____________
☐ Loss of function of an arm or leg ____________
☐ Disfiguring scars ____________

The doctor has explained to me the most likely complications or undesired results that might occur in this operation or medical procedure and I understand them. The doctor has offered to detail the less likely complications of [sic] undesired results which, even if rare, could occur.

______ I do ______ I do not wish to have a full description of all the possible complications given to me.

I hereby authorize and direct the above named physician with associates or assistants to provide such additional services as they might deem reasonable and necessary including, but not limited to, the administration of any anesthetic agent, or the services of the X-ray department or laboratories, and I hereby consent thereto.

I hereby state that I have read and understand this consent and that all blanks were filled in prior to my signature.

Date: ____________ Time ____________ a.m. p.m.

Signature of Patient ____________
Signature of Relative (where required) ____________
Witness ____________

I certify that I have personally completed all blanks in this form and explained them to the patient or his representative before requesting the patient or his representative to sign it.

________________________ (signature of named physician)[12]

This form seems to strike a balance between requiring physicians to prepare unique forms for each procedure and using a standard form for all procedures. Here a standard form is used, but each case is treated uniquely, based on the manner in which the blanks are filled in. The danger exists, however, that providing this time-saving device to the physician may detract from the goal of promoting individual autonomy and encouraging rational decision making. The more routinized and standardized the procedure becomes, the less likely it will be that these goals will be attained. Informed consent is a *process* of giving and receiving information. The consent *form,* on the other hand, is merely physical evidence that the process has taken place. Its traditional role has been to help protect the physician in case the patient later denies that informed consent was in fact obtained.

---

[12] Ohio St. sec. 2317.54(D).

It would be a tragedy if the promotion of a uniform consent form led to making the process itself just one step that both physician and patient come to regard as unnecessary red tape, and, as a consequence, they fail to engage in the dialogue envisioned by the doctrine. Each procedure has its own unique risks, and each patient is in some manner unique. Only by going through the difficult task of tailoring information to fit the procedure and the patient is it likely that either individual autonomy or rational decision making will be promoted. If neither of these functions is furthered, the process becomes meaningless.

CONCLUSIONS

While few conclusions can be drawn from this group of statutes, one can see a trend toward treating the written consent form as a complete defense in an informed consent case (barring fraud or misrepresentation). Most legislatures that have acted on this question to date are concerned with *disclosure* on the part of the physician, rather than *understanding* on the part of the patient. In this approach they are consistent with the appellate courts. Some legislatures are less consistent in permitting the physician to limit disclosures based on "standard medical practice" rather than on the patient's need to know. In this regard, many of the statutes must be seen as anti-self-determination measures.

Ann. Rev. Med. 1978. 29:15–22
Copyright © 1978 by Annual Reviews Inc. All rights reserved

POLYMYALGIA RHEUMATICA AND GIANT CELL ARTERITIS

◆7261

Robert E. Ettlinger, M.D., Gene G. Hunder, M.D., and L. Emmerson Ward, M.D.

Division of Rheumatology, Department of Internal Medicine, Mayo Clinic
and Mayo Foundation, Rochester, Minnesota 55901

Growing appreciation of the close relationship between polymyalgia rheumatica (PMR) and giant cell arteritis (GCA) demands that the two conditions be reviewed conjointly. Speculation on the nature of their kinship has ranged from the opinion that GCA is the underlying cause of most cases of PMR to the belief that PMR is a separate process that may coexist with GCA. In this review, we summarize the characteristics of each disease and examine the evidence that tends to unite or divide them as entities.

DEFINITION OF PMR AND GCA

As a minimum, most authors (1–3) would give a diagnosis of PMR to a middle-aged or older patient with complaints of proximal musculoskeletal discomfort lasting several weeks or longer and accompanied by constitutional symptoms such as fever, malaise, and weight loss. Although exceptions have been reported (4), the erythrocyte sedimentation is so commonly accelerated that some authors (5, 6) consider its increase as an additional criterion for diagnosis.

In most cases of PMR no other illness is found, even by diligent search and prolonged observation. When an associated condition can be identified, GCA is by far the most common. GCA is a vasculitis, chiefly of the large and medium-sized arteries of the body. The inflammatory cells within the vessel wall consist of lymphocytes, macrophages, and plasma cells—sometimes limited to the elastic laminae or adventitia, but more often spread throughout all the layers. The intimal layer displays fibrous proliferation and the elastic laminae are disrupted and frayed. Giant cells are often present at the intima-medial junction and may contain ingested elastic fibers. Fibrinoid necrosis is notably absent.

Histologic findings indistinguishable from GCA can be observed in Takayasu's disease (7). However, differences of clinical manifestations, age, and affected population distinguish this disease, giant cell arteritis of the central nervous system (8), and "juvenile temporal arteritis" (9) from GCA.

0066-4219/78/0401-0015$01.00

CLINICAL FEATURES

Polymyalgia Rheumatica

The diagnosis of PMR is usually reserved for patients aged 50 years or more. The mean age at onset is 65 years, and the female-to-male incidence ratio is 2:1 (2, 5, 10). PMR appears to be fairly common, at least in Europe and the United States, where rather large series have been reported.

Most patients have been in good health before developing PMR. The myalgias and arthralgias usually evolve insidiously over several weeks or months. In two thirds of the cases, the periarticular areas of the neck and shoulder girdle are the first to become painful; in the remainder, the hips and thighs are involved at the onset. The discomfort may begin in one shoulder or hip region, but usually becomes bilateral within weeks and may involve most of the proximal and axial musculature and tendinous attachments. Morning stiffness or "gelling," like that characteristic of rheumatoid arthritis, is usually present. The discomfort is exacerbated by even minimal motion of affected regions. Muscular strength is unimpaired objectively, although the pain involved in testing may not allow the patient to cooperate maximally.

Pain is not usually referred specifically to the joints themselves, although limitations of motion of the shoulders and hips may be noted (5). Some of this limitation may be due to pain alone, but the presence of synovitis in PMR is being appreciated increasingly and undoubtedly is the cause of some of the findings in the process. Synovitis has been documented by biopsy (3, 4), synovianalysis (11), and joint scintigram (12).

In contrast, muscle biopsies in PMR are usually reported as negative, although occasional minor inflammatory infiltrates are seen in muscular septae or perivascular areas (13) and nonspecific type II muscle fiber atrophy is found (14).

Both low-grade fever and weight loss are present in the majority of cases and malaise, fatigue, and depression may have been experienced for months when the diagnosis is first suspected (2).

Giant Cell Arteritis

The population susceptible to GCA appears to be the same as that susceptible to PMR (10). A recent study has revealed an average annual incidence of 11.7 per 100,000 persons aged 50 and older, and a prevalence (active and inactive disease included) of 133 per 100,000 persons aged 50 and more (K. A. Huston and associates, unpublished observations).

Headache occurs in almost half of the patients with GCA. It may be localized to the temporal or occipital regions, but often is nonspecific in character. Involvement of the cranial arteries may result in scalp pain with combing of the hair, wearing eyeglasses, or simply placing the head on a pillow. The temporal artery may be tender, thickened, nodular, or pulseless. Involvement of other cranial arteries may produce the pathognomonic symptoms of jaw claudication and the lingual Raynaud's phenomenon.

The most feared complication of GCA, however, is blindness. It may appear abruptly, but in most cases is preceded by several days or weeks of episodic visual loss or headaches. Blindness is generally caused by ischemic optic neuritis resulting from occlusion of the posterior ciliary and ophthalmic arteries. Of untreated patients, as many as half have at least transient visual loss; in a third, a permanent visual loss will develop (15). If blindness occurs in one eye, the other will be similarly affected within days or weeks in 25% of cases unless treatment is given (16).

GCA involvement of large arteries such as the carotids may produce hemiplegia and cranial nerve palsies; and when the vertebral and basilar systems are involved, various neurologic symptoms—including seizures and vertigo—may result (5, 16). Both Bruk (4) and Hamrin (5) detected bruits in more than half of their patients with arteritis, most commonly over the axillary, brachial, carotid, and subclavian arteries. However, Klein et al (17) found only 9% of a large series of GCA patients had symptoms attributable to large-artery involvement. Intermittent claudication of the upper extremities was most common. Other reported consequences of GCA include myocardial infarction, coronary insufficiency, renal artery stenosis, and mesenteric infarctions.

PMR may either herald the presence of GCA or appear when the arteritic symptoms are well established or even subsiding. Half to three quarters of all patients with GCA have PMR at some time during their illness (2, 5, 10).

RELATIONSHIP BETWEEN PMR AND GCA

Paulley & Hughes (18, 19) first suggested that the symptoms of PMR might be the expression of an underlying arteritis which was not always clinically evident. This observation was supported by Alestig & Barr (7) who found temporal arteritis in PMR patients without symptoms or signs of arteritis. When interpreting subsequent reports in which the incidence of GCA in PMR is re-examined (3–6, 11), one must bear in mind that biopsy of a temporal artery may easily miss inflammation because of irregular distribution; that arteritis may affect other vessels, whether present in the temporal artery or not; and that patient selection may have much influence on determinations of incidence if patients with symptoms or signs suggestive of arteritis are biopsied preferentially. Bruk (4) reported 15 positive specimens among 33 from patients with clinical evidence of arteritis, but none positive among 15 from PMR patients with minimal or no clinical evidence of arteritis. A similar clinical distinction in the incidence of arteritis was noted by Kattwinkel & Fernandez-Herlihy (6) and by Wilske & Healey (11). A higher incidence of "occult" GCA—40%—in patients with PMR was noted more recently by Fauchald, Rygvold & Øystese (10) and by Hamrin (5). In summary, the incidence of GCA among all patients with PMR may approach 50%. Depending on how strictly one can exclude those patients with either symptoms or signs of arteritis, the incidence of arteritis in the remainder of PMR cases may be as low as 15%.

The frequent coexistence of GCA and PMR has led several investigators to conclude that the proximal muscular discomfort in PMR is due to an underlying vasculitis (5, 7, 10, 11), despite the lack of histologic evidence for vasculitis in many cases. However, it is our current view that, in GCA and PMR, there may be a common etiologic event to which two pathologic responses are possible—namely, arteritis or synovitis. The varied clinical and pathologic manifestations in a patient with either or both conditions would then depend upon the proportional mixture of these two responses.

LABORATORY EVALUATION

The erythrocyte sedimentation is nearly always abnormally rapid in PMR or GCA, and often strikingly rapid. In a series of 93 cases of PMR with and without GCA (5), the erythrocyte sedimentation rate ranged from 47–155 mm in 1 hr (Westergren) with a mean of 103 mm. With clinical improvement, the erythrocyte sedimentation slows; a return of rapid fall may signal recurrence. Despite complete remission of the signs and symptoms of earlier GCA or PMR, the erythrocyte sedimentation may remain somewhat higher than normal, however.

A mild normochromic or hypochromic anemia, commonly present in both GCA and PMR, returns to normal with clinical improvement (11). The bone marrow findings are nondiagnostic (2). Leukocytes and platelets do not show consistent abnormalities.

The serum protein electrophoretic pattern in GCA and PMR usually is characterized by decreased albumin and increased α-2 fractions. As with the erythrocyte sedimentation and hemoglobin, correction of this abnormality parallels clinical improvement.

Tests for rheumatoid factor, antinuclear antibody, and L.E. cells are no more frequently positive than expected in an elderly population. The electromyographic and serum enzyme findings that would indicate muscle injury are normal, and no erosive changes are consistently noted on radiographs of joints.

Arteriograms of the major vessels can be helpful in GCA by documenting the presence and distribution of large-artery involvement (17). Superficial temporal artery arteriography may be advantageous in directing the surgeon's attention to positive areas prior to biopsy; however, to obtain a lengthy specimen is easy enough so that we do not routinely undertake the delay and expense of preliminary search for precise locations of involvement.

ETIOLOGIC INVESTIGATIONS

No etiologic agent has been identified in PMR or GCA. A genetic predisposition to their development is suggested by the presence of these conditions in a reported total of 10 pairs of first-degree relatives (20); but to date, HLA haplotype determinations have not supported this implication (21, 22).

Recently, attention has focused on the possibility that alteration of the immune system has a part in the pathogenesis of these diseases. Since the internal elastic membrane is fragmented and surrounded by an inflammatory exudate with giant cells often containing bits of ingested elastica, some have suspected that the membrane contains the autoantigen in GCA (23, 24). Evidence for the participation of the immune system in PMR and GCA comes from studies showing an increased number of circulating immunoblasts in PMR (25), the presence of anti-IgG activity in temporal artery biopsy specimens (26), and from immunofluorescent investigations showing immunoglobulin deposition within the wall of temporal arteries involved with GCA (23). Cell-mediated mechanisms are suggested by the work of Hazleman, MacLennan & Esiri (24), who found DNA synthesis in peripheral blood lymphocytes from patients with PMR and GCA to be stimulated by homogenates of human artery and muscle.

DIFFERENTIAL DIAGNOSIS

Although in most cases of PMR either GCA or no underlying illness will be discovered, occasionally there is an associated infection or neoplasm (3). If the history and physical examinations do not suggest either of the latter, however, little is to be gained by a more extensive radiographic or laboratory search. The early stages of rheumatoid arthritis, ankylosing spondylitis, systemic lupus erythematosus, or systemic sclerosis may be indistinguishable from PMR and a period of observation for the more characteristic features of these illnesses may be required for differentiation. Usually, however, the presence of other features of these diseases will help in making the proper diagnosis. In Myles' (27) series of 84 cases with initial diagnosis of PMR, long-term follow-up showed only 7 to be incorrect: 5 were seronegative polyarthritis, 1 was ankylosing spondylitis, and 1 remained unclassified.

After the generalized constitutional symptoms have been recorded and the pattern of musculoskeletal discomfort defined, the physician should inquire about symptoms suggestive of arterial involvement (GCA). Careful attention should be directed to the cranial and peripheral arteries. A funduscopic examination should be performed and the completeness of the visual fields checked. Confirmatory laboratory studies should include determinations of hemoglobin, erythrocyte sedimentation rate, and serum protein electrophoresis.

If PMR is present and if no underlying infectious or neoplastic process is evident, temporal artery biopsy should be carried out in most cases. Some patients who have no signs or symptoms of arterial involvement, or who have had mild symptoms of PMR for a long or a very short period and have only moderate laboratory manifestations of systemic inflammation, may be managed without a biopsy—provided that close follow-up can be maintained and that the patient can recognize and report early arteritic symptoms. Such patients have a low incidence of arteritic involvement (4, 10) and are unlikely to develop blindness (4). Unlike some authors (2, 10, 11), but in agreement with others (6, 27), we feel that not all PMR patients need to be treated with corticosteroids; some may get adequate relief from their musculoskeletal symptoms with aspirin, indomethacin, or other nonsteroidal, anti-inflammatory drugs.

Biopsy of the temporal arteries should be performed if corticosteroids are to be included in treatment, if there are signs or symptoms that suggest arterial inflammation, or if worsening, lack of improvement, or relapse occurs in a patient not biopsied previously. A clinically involved portion of artery should be biopsied if possible. When there are no findings referable to the arteries, a generous specimen (4–6 cm) of one temporal artery should be obtained and examined histologically at multiple levels (28). Since GCA is unilateral in 5% of cases (28), the contralateral artery should be biopsied if frozen sections reveal no disease in the first.

TREATMENT AND RESPONSE

Rapid improvement in the musculoskeletal symptoms of PMR can be expected within 12–24 hr after corticosteroid therapy begins. The arteritic manifestions of GCA, such as claudication of the extremities and scalp pain, also respond (3, 17) —although visual loss, when complete, is not usually reversed. So impressive is the relief from the musculoskeletal discomfort that some authors have advocated the use of small doses of corticosteroids as a confirmation of the diagnosis of PMR. Although we agree that a prompt response is consistent with the diagnosis, we feel that such a trial is rarely necessary and is not specific anyway. Similar responses of the myalgias and morning stiffness of early rheumatoid arthritis and systemic lupus erythematosus may be obtained just as readily with this therapy.

Many authors advocate treating all PMR patients with steroids, since they believe that the symptoms in all cases represent underlying GCA. As previously explained, however, we feel that in many patients with mild manifestations and no evidence of arteritis the discomfort is due primarily to a synovitis that can be managed satisfactorily with nonsteroidal anti-inflammatory medications. In such cases, a physical therapy program including heat and exercises may be helpful. If these measures are not effective, corticosteroids can be considered.

All patients with biopsy evidence of arteritis should be treated with corticosteroids. Initial dosage of 40–60 mg of prednisone per day will provide nearly universal improvement of symptoms and protect the patient from blindness. Alternate-day dosage of 90 mg prednisone, though devoid of most of the side effects of hypercortisonism, usually is not effective in suppressing the arteritis (29).

Close follow-up is essential in management of PMR and GCA, and a monthly communication with the patient through the early part of the course is recommended. Once the clinical features of the illness have subsided and the laboratory abnormalities reflecting systemic inflammation have returned to normal, attempts should be made to taper the steroid dosage to the lowest necessary maintenance level. Reductions of about 10% of the total daily dose may be made at weekly intervals initially. Although erythrocyte sedimentation may increase slightly as tapering proceeds, this does not signal a relapse if symptoms do not return. If a recrudescence is suspected, tapering may be suspended temporarily or smaller reductions can be made at longer intervals. If a relapse occurs, doses that adequately suppressed the disease previously should be resumed. Whereas freedom from musculoskeletal symptoms can usually be maintained at a dosage of 5–15 mg per day, higher doses may be necessary to prevent arterial complications. When the dosage is less than 10 mg per day, reductions should be made in steps of 1 mg per month if tolerated.

Disease Duration and Relapse

Analysis of the clinical course of patients who had both PMR and GCA prior to the introduction of corticosteroids (1) revealed that although spontaneous remission can occur in as little time as 6 months, the illness usually persists longer —in many instances for 2 or 3 years or more (2, 10, 27, 30). There is no evidence that corticosteroid treatment shortens the duration at all (2, 10, 27, 30). It is reported that if treatment is discontinued in less than 2 years, relapse occurs in about 20% of cases of GCA or PMR, usually within 1 year of steroid withdrawal (10). Among patients with GCA described by Beevers, Harpur & Turk (30), most relapses occurred during the first year of treatment, while the patients were on an average prednisone dose of 11.4 mg and had a mean erythrocyte sedimentation of 53 mm, which suggests too rapid steroid withdrawal.

CONCLUSION

PMR is a clinical syndrome characterized by pain and stiffness in the region of the axial skeleton, nonspecific systemic symptoms, and increased erythrocyte sedimentation, generally in patients past the age of 50. GCA is characterized by a granulomatous inflammatory reaction, usually including giant cells, which occurs chiefly in large and medium-sized arteries. Various manifestations may result from the arteritis, including headache, claudication of the jaw, lingual Raynaud's phenomenon, arterial bruits, and visual loss secondary to ischemic optic neuritis. In addition, PMR occurs in half or more of patients who have GCA.

Many patients who have PMR—perhaps 50%—can be shown to have underlying GCA. Whether the remainder also have arteritis that is not demonstrable by available diagnostic methods remains in dispute.

Patients with GCA, or those suspected of having it, should be treated with doses of corticosteroids adequate to suppress the inflammatory reaction in the arteries, particularly to prevent loss of vision and other vascular complications.

Literature Cited

1. Barber, H. S. 1957. Myalgic syndrome with constitutional effects: polymyalgia rheumatica. *Ann. Rheum. Dis.* 16:230–37
2. Hamilton, C. R. Jr., Shelley, W. M., Tumulty, P. A. 1971. Giant cell arteritis: including temporal arteritis and polymyalgia rheumatica. *Medicine* 50:1–27
3. Hunder, G. G., Disney, T. F., Ward, L. E. 1969. Polymyalgia rheumatica. *Mayo Clin. Proc.* 44:849–75
4. Bruk, M. I. 1967. Articular and vascular manifestations of polymyalgia rheumatica. *Ann. Rheum. Dis.* 26:103–16
5. Hamrin, B. 1972. Polymyalgia arteritica. *Acta Med. Scand. Suppl.* 533:1–131
6. Kattwinkel, N., Fernandez-Herlihy, L. 1970. Polymyalgia rheumatica

P 24839

COLLEGE OF MEDICINE LIBRARY C. U.

and temporal arteritis. *Lahey Clin. Found. Bull.* 19:41–48
7. Alestig, K., Barr, J. 1963. Giant-cell arteritis: a biopsy study of polymyalgia rheumatica, including one case of Takayasu's disease. *Lancet* 1:1228–30
8. Nurick, S., Blackwood, W., Mair, W.G.P. 1972. Giant cell granulomatous angiitis of the central nervous system. *Brain* 95:133–42
9. Lie, J. T., Gordon, L. P., Titus, J. L. 1975. Juvenile temporal arteritis: biopsy study of four cases. *J. Am. Med. Assoc.* 234:496–99
10. Fauchald, P., Rygvold, O., Øystese, B. 1972. Temporal arteritis and polymyalgia rheumatica: clinical and biopsy findings. *Ann. Intern. Med.* 77:845–52
11. Wilske, K. R., Healey, L. A. 1967. Polymyalgia rheumatica: a manifestation of systemic giant-cell arteritis. *Ann. Intern. Med.* 66:77–86
12. O'Duffy, J. D., Wahner, H. W., Hunder, G. G. 1976. Joint imaging in polymyalgia rheumatica. *Mayo Clin. Proc.* 51:519–24
13. Gordon, I., Rennie, A. M., Branwood, A. W. 1964. Polymyalgia rheumatica: biopsy studies. *Ann. Rheum. Dis.* 23:447–55
14. Brooke, M. H., Kaplan, H. 1972. Muscle pathology in rheumatoid arthritis, polymyalgia rheumatica, and polymyositis. A histochemical study. *Arch. Pathol.* 94:101–18
15. Ladas, G. J. 1973. Temporal arteritis: a review of the Washington Hospital Center experience with comparison of the literature. *Med. Ann. DC* 42:273–76
16. Hollenhorst, R. W., Brown, J. R., Wagener, H. P., Shick, R. M. 1960. Neurologic aspects in temporal arteritis. *Neurology* 10:490–98
17. Klein, R. G., Hunder, G. G., Stanson, A. W., Sheps, S. G. 1975. Large artery involvement in giant cell (temporal) arteritis. *Ann. Intern. Med.* 83:806–12
18. Paulley, J. W. 1956. Anarthritic rheumatoid disease (Letter to the Editor). *Lancet* 2:946
19. Paulley, J. W., Hughes, J. P. 1960. Giant-cell arteritis, or arteritis of the aged. *Br. Med. J.* 2:1562–67
20. Liang, G. C., Simkin, P. A., Hunder, G. G., Wilske, K. R., Healey, L. A. 1974. Familial aggregation of polymyalgia rheumatica and giant cell arteritis. *Arthritis Rheum.* 17:19–24
21. Terasaki, P. I., Healey, L. A., Wilske, K. R. 1976. Distribution of HLA haplotypes in polymyalgia rheumatica (Letter to the Editor). *N. Engl. J. Med.* 295:905
22. Hunder, G. G., Taswell, H. F., Pineda, A. A., Elveback, L. R. 1977. HLA antigens in patients with giant cell arteritis and polymyalgia rheumatica. *J. Rheumatol.* In press
23. Liang, G. C., Simkin, P. A., Mannik, M. 1974. Immunoglobulins in temporal arteries: an immunofluorescent study. *Ann. Intern. Med.* 81:19–24
24. Hazleman, B. L., MacLennan, I.C.M., Esiri, M. M. 1975. Lymphocyte proliferation to artery antigen as a positive diagnostic test in polymyalgia rheumatica. *Ann. Rheum. Dis.* 34:122–27
25. Eghtedari, A. A., Esselinckx, W., Bacon, P. A. 1976. Circulating immunoblasts in polymyalgia rheumatica. *Ann. Rheum. Dis.* 35:158–62
26. Waaler, E., Tönder, O., Milde, E. J. 1976. Immunological and histological studies of temporal arteries from patients with temporal arteritis and/or polymyalgia rheumatica. *Acta Pathol. Microbiol. Scand. Sect. A* 84A:55–63
27. Myles, A. B. 1975. Polymyalgia rheumatica and giant cell arteritis: a seven-year survey. *Rheumatol. Rehabil.* 14:231–35
28. Klein, R. G., Campbell, R. J., Hunder, G. G., Carney, J. A. 1976. Skip lesions in temporal arteritis. *Mayo Clin. Proc.* 51:504–10
29. Hunder, G. G., Sheps, S. G., Allen, G. L., Joyce, J. W. 1975. Daily and alternate-day corticosteroid regimens in treatment of giant cell arteritis: comparison in a prospective study. *Ann. Intern. Med.* 82:613–18
30. Beevers, D. G., Harpur, J. E., Turk, K. A. D. 1973. Giant cell arteritis—the need for prolonged treatment. *J. Chronic Dis.* 26:571–84

Ann. Rev. Med. 1978. 29:23–29
Copyright © 1978 by Annual Reviews Inc. All rights reserved

PROSTAGLANDINS AND HYPERCALCEMIC STATES

◆7262

Hannsjörg W. Seyberth, M.D.
Klinikum der Universität Heidelberg, Kinderklinik, Im Neuenheimer Feld 150,
D 6900 Heidelberg, Germany

Lawrence G. Raisz, M.D.
Department of Medicine, University of Connecticut Health Center, Farmington,
Connecticut 06032

John A. Oates, M.D.
Departments of Medicine and Pharmacology, Vanderbilt University, Nashville,
Tennessee 37232

INTRODUCTION

In 1970, Klein & Raisz (1) reported that prostaglandins of the E series (PGE) were as potent stimulators of bone resorption in organ cultures as parathyroid hormone (PTH) and Chase & Aurbach (2) showed that PGE also shared with PTH the ability to raise cyclic adenosine 3′5′-monophosphate (cAMP) content in bone. Subsequent studies of the effects of prostaglandin on calcium regulation and bone metabolism have recently been reviewed (3). These compounds have now been implicated as mediators of pathologic bone resorption in rheumatoid arthritis (4), periodontal disease (5, 6), and malignancy (7–9).

The major purpose of this chapter is to summarize those observations that support the concept that prostaglandins are involved in the pathogenesis of hypercalcemia in malignancy in both experimental animals and man. These observations have important implications for the diagnosis and management of this common complication of neoplastic disease.

EFFECTS OF PROSTAGLANDINS ON BONE TISSUE

Although prostaglandins of the E series are the most potent stimulators of bone resorption in vitro, all of the classical primary prostaglandins so far tested, and a number of their metabolites, can also stimulate resorption to greater or lesser

0066-4219/78/0401-0023$01.00

degrees (7, 10–12). The more recently discovered unstable products of prostaglandin cyclo-oxygenase, such as thromboxane A_2 (13) and prostacyclin (14, 15), have not been studied in bone culture. However, the endoperoxide precursors PGG_2 and PGH_2 have been shown to produce a transient calcium release from bone with little prolonged resorptive response (12). The effects of PGE_2 on bone generally resemble those of PTH. Both compounds increase the size and number of osteoclasts as well as the area of ruffled border (M. E. Holtrop and L. G. Raisz, unpublished observations). Both compounds produce an early increase in the release of lysosomal enzymes from bone (G. Eilon and L. G. Raisz, unpublished observations). PGE_2 can also inhibit bone collagen synthesis (16), but higher concentrations than those that stimulate bone resorption are required. The fact that PTH and PGE_2 are not synergistic at low concentrations and that no further resorption is obtained when the two compounds are combined at maximal concentrations indicates that they act through a common cellular pathway (1, 3). However, there is evidence that the receptors for PTH and PGE in bone cells are different (17–19).

PROSTAGLANDINS IN ANIMAL MODELS OF TUMOR-ASSOCIATED HYPERCALCEMIA

The concept that prostaglandins might mediate the hypercalcemia of malignancy was first presented by Tashjian and his associates (7) in 1972 using a transplantable mouse fibrosarcoma $HSDM_1$. Subsequently, these investigators extended their studies to the VX_2 squamous-cell carcinoma in the rabbit (8). These tumors could produce hypercalcemia without bone invasion. In the rabbit model the hypercalcemia was associated with an increase in osteoclasts in sites distant from the tumor (20). Cells from both tumors secreted a bone-resorbing factor in culture that appeared to be attributable to PGE_2 as measured by radioimmunoassay (7, 8). By mass spectrometric analysis, homogenates and minced tissue of VX_2 carcinoma were shown to convert both exogenous and endogenous arachidonic acid, the major natural substrate for the prostaglandin cyclo-oxygenase, to PGE_2 (21). Treatment of the cultured cells with indomethacin, a potent inhibitor of prostaglandin cyclo-oxygenase (22), blocked the production of radioimmunoassayable PGE_2 and the release of bone-resorbing factor in vitro (7, 8). Moreover, treatment of hypercalcemic animals with indomethacin generally lowered the elevated serum calcium concentration (7, 8, 21). Finally, the concentration of immunoreactive PGE_2 in the blood collected from the femoral vein of the tumor-bearing leg was higher than that of the opposite femoral vein in the rabbits (8).

While the observations listed above support the hypothesis that excessive synthesis and secretion of PGE_2 by the tumor was the cause of hypercalcemia (23), certain problems and alternative possibilities have to be considered. Tumors clearly could produce prostaglandins in vitro, but non-neoplastic tissue can also produce prostaglandins. Bone itself can produce prostaglandins in organ culture in the presence of complement-sufficient heterologous serum (24) and when treated with crude collagenase (25). Another problem is that PGE_2 secreted by a tumor in vivo would pass through the lung where it is likely to be almost completely metabolized (26–29). Moreover, the exogenous administration of PGE_2 has usually not produced hypercalcemia in experimental animals (1, 30, 31) although a small increase was observed when large doses of PGE_2 were infused into rats (32). Finally, in rabbits with VX_2 carcinoma the plasma concentration of radioimmunoassayable PGE_2 was not elevated at the time of onset of hypercalcemia (8) although mass spectrometric measurements of the principle plasma metabolite of PGE_2, 15-keto-13,14-dihydroxy-PGE_2 (15-keto-H_2-PGE_2), did show an increase prior to the increase in serum calcium concentration (21, 33). This observation has been corroborated using a radioimmunoassay for 15-keto-H_2-PGE_2 in mice bearing the $HSDM_1$ fibrosarcoma (34) and in the VX_2 carcinoma-bearing rabbits (35). Increased PGE_2 production was also reflected in an increase in the urinary excretion of 7-α-hydroxy-5,11-diketotetranorprostane-1,16-dioic acid (PGE-M), a major urinary metabolite of the E prostaglandins (21). These findings indicate that determination of prostaglandin metabolites might be more sensitive than measurements of the parent compound for detection of changes in biosynthesis. Moreover, the plasma metabolites do retain some biological activity. In particular, the 13,14-dihydro-metabolites of PGE are quite active as stimulators of bone resorption (11, 12) and also effective in increasing cAMP concentration in rat osteogenic sarcoma cells (36). Even 15-keto-H_2-PGE_2, which is biologically much less active than the parent compound, might function as a mediator of hypercalcemia since it is present in such high concentrations (21). Further studies of type and localization of prostaglandins in tumors, in the circulation, and in the affected bone are needed before the mechanism of hypercalcemia can be determined precisely.

PROSTAGLANDINS IN HUMAN HYPERCALCEMIA OF MALIGNANCY

The first evidence that prostaglandins might mediate hypercalcemia in human malignancy was provided by a case report of a patient with renal-cell adenocarcinoma whose hypercalcemia was improved by treatment with indomethacin (37). Although indomethacin is not completely specific as an inhibitor of prostaglandin cyclo-oxygenase, the response does suggest that the hypercalcemia was mediated by prostaglandins. Indomethacin generally does not reduce hypercalcemia or block bone resorption produced by other agents, such as vitamin D metabolites or parathyroid hormone (24, 25, 38, 39). In 1975, Robertson and his co-workers (40) described a patient with renal-cell carcinoma in which hypercalcemia was associated with an elevated immunoreactive PGE concentration in the blood, but low plasma PTH. Since that time a number of reports have described either increased prostaglandin concentrations or indomethacin responsiveness in hypercalcemia due to various malignant diseases (9, 38, 39, 41–45). On the other hand, many patients do not respond to indomethacin so alternative mechanisms for hypercalcemia, such as ectopic parathyroid hormone secretion, must exist (23, 46, 47). In hematologic neoplasms, such as multiple myeloma, Mundy and associates (48) have obtained evidence for the production of an osteoclast-activating factor chemically unrelated to prostaglandins or PTH.

In a series of patients with solid tumors, predominantly squamous-cell and undifferentiated carcinomas of lung, pancreas, and uterine cervix, quantitative analysis of the major urinary metabolite of the E prostaglandins, PGE-M, was used to characterize the subset of hypercalcemic patients who showed prostaglandin overproduction (9, 38). In these patients the median PGE-M excretion was 59 ng per mg creatinine compared with 7 ng per mg for controls. PGE-M excretion was slightly elevated in normocalcemic patients with similar types of solid tumors. PGE-M excretion was not increased in hypercalcemic patients with primary hyperparathyroidism or those with hematologic neoplasms, such as multiple myeloma, malignant lymphoma, or lymphosarcoma. The latter observation supports the concept that hematologic neoplasms produce a different osteoclast-activating factor (48). In the patients with high PGE-M excretion, indomethacin at 75–150 mg per day or aspirin at 1.8–4.8 g per day decreased PGE-M excretion (presumably by inhibiting prostaglandin cyclo-oxygenase), reduced serum calcium concentration, and decreased urinary calcium excretion. The serum calcium was brought to normal in patients without bone metastases but was only partially reduced in those patients with bone metastases. These drugs had no effect on serum calcium in patients whose hypercalcemia was not associated with elevated PGE-M excretion. In patients with solid tumors whose hypercalcemia did not appear to be associated with increased PGE-M levels, there was a tendency for the renal excretion of cyclic AMP and serum chloride concentration to be increased, changes that are also seen in primary hyperparathyroidism. However, in all of the patients with hypercalcemia in this series, immunoreactive PTH was undetectable irrespective of PGE-M excretion.

Patients with breast cancer did not show grossly increased excretion of PGE-M although the values are somewhat higher in hypercalcemic than in normocalcemic patients (H. W. Seyberth, unpublished observations). This finding and the failure of indomethacin to reduce hypercalcemia in patients with breast cancer (47) suggest that prostaglandins are not important mediators. On the other hand, breast cancer tissue has been shown to release bone-resorbing factors in culture that can in part be inhibited by aspirin (49). Such a factor could be a prostaglandin but its relevance to in vivo production by these tumors is unknown.

CONCLUSION

There is considerable circumstantial evidence to support the concept that prostaglandins of the E series can mediate hypercalcemia in neoplastic disease. However, other metabolites of prostaglandin endoperoxides could contribute to the hypercalcemia and the site of the production of prostaglandins is not established. Further analysis of prostaglandin endoperoxide metabolism in tumor and bone tissue and in the circulation will clarify the specific mechanism. The development of quantitative analysis of prostaglandin metabolites has made it possible to identify prostaglandin-dependent hypercalcemia. Therapeutic application of the inhibitors of prostaglandin cyclo-oxygenase in the treatment of patients with prostaglandin-mediated hypercalcemia is feasible, but its general usefulness remains to be evaluated. The development of antagonists to the action of prostaglandin at the affected end organ site, presumably bone, would represent an important therapeutic advance.

Literature Cited

1. Klein, D. C., Raisz, L. G. 1970. Prostaglandins: Stimulation of bone resorption in tissue culture. *Endocrinology* 86:1436–40
2. Chase, L. R., Aurbach, G. D. 1970. The effect of parathyroid hormone on the concentration of adenosine 3′5′-monophosphate in skeletal tissue *in vitro*. *J. Biol. Chem.* 245:1520–26
3. Dietrich, J. W., Raisz, L. G. 1975. Prostaglandin in calcium and bone metabolism. *Clin. Orthop. Relat. Res.* 111:228–37
4. Robinson, D. R., Tashjian, A. H. Jr., Levine, L. 1975. Prostaglandin-stimulated bone resorption by rheumatoid synovia: A possible mechanism for bone destruction in rheumatoid arthritis. *J. Clin. Invest.* 56:1181–88
5. Goodson, J. M., Dewhirst, F. E., Brunetti, A. 1974 Prostaglandin E_2 levels and human periodontal disease. *Prostaglandins* 6:81–85
6. Gomes, B. C., Hausmann, E., Weinfeld, N., De Luca, C. 1976. Prostaglandins: Bone resorption stimulating factors released from monkey gingiva. *Calcif. Tissue Res.* 19:285–93
7. Tashjian, A. H. Jr., Voelkel, E. F., Levine, L., Goldhaber, P. 1972. Evidence that bone resorption-stimulating factor produced by mouse fibrosarcoma cells is prostaglandin E_2: A new model for the hypercalcemia of cancer. *J. Exp. Med.* 136:1329–43
8. Voelkel, E. F., Tashjian, A. H. Jr., Franklin, R., Wasserman, E., Levine, L. 1975. Hypercalcemia and tumor-prostaglandins: The VX_2 carcinoma model in the rabbit. *Metabolism* 24:973–86
9. Seyberth, H. W., Segre, G. V., Morgan, J. L., Sweetman, B. J., Potts, J. T. Jr., Oates, J. A. 1975. Prostaglandins as mediators of hypercalcemia associated with certain types of cancer. *N. Engl. J. Med.* 293:1278–83
10. Dietrich, J. W., Goodson, J. M., Raisz, L. G. 1975. Stimulation of bone resorption by various prostaglandins in organ culture. *Prostaglandins* 10:231–40
11. Tashjian, A. H. Jr., Tice, J. E., Sides, K. 1977. Biological activities of prostaglandin analogues and metabolites on bone in organ culture. *Nature* 266:645–47
12. Raisz, L. G., Dietrich, J. W., Simmons, H. A., Seyberth, H. W., Hubbard, W., Oates, J. A. 1977. Effect of prostaglandin endoperoxides and metabolites on bone resorption *in vitro*. *Nature* 267:532–34
13. Hamberg, M., Svensson, J., Samuelsson, B. 1975. Thromboxanes: A new group of biologically active compounds derived from prostaglandin endoperoxides. *Proc. Natl. Acad. Sci. USA* 72:2994–98
14. Moncada, S., Gryglewski, R., Bunting, S., Vane, J. R. 1976. An enzyme isolated from arteries transforms prostaglandin endoperoxides to an unstable substance that inhibits platelet aggregation. *Nature* 263:663–65
15. Johnson, R. A., Morton, D. R., Kinner, J. H., Gorman, R. R., McGuire, J. C., Sun, F. F., Whittaker, N., Bunting, S., Salmon, J., Moncada, S., Vane, J. R. 1976. The chemical structure of prostaglandin X (prostacyclin). *Prostaglandins* 12:915–28

16. Raisz, L. G., Koolemans-Beynen, A. R. 1974. Inhibition of bone collagen synthesis by prostaglandin E_2 in organ culture. *Prostaglandins* 10:377–85
17. Yu, J. H., Wells, H., Ryan, W. J. Jr., Lloyd, W. S. 1976. Effects of prostaglandins and other drugs on the cyclic AMP content of cultured bone cells. *Prostaglandins* 12:501–13
18. Chase, L. R., Obert, K. A. 1975. Selective proteolysis of the receptor for parathyroid hormone in skeletal tissue. *Metabolism* 24:1067–71
19. Rao, L. G., Ng, B., Brunette, D. M., Heersche, J.N.M. 1977. Parathyroid hormone- and prostaglandin E_1-response in a selected population of bone cells after repeated subculture and storage at −80 C. *Endocrinology* 100:1233–41
20. Hough, A. Jr., Seyberth, H., Oates, J., Hartmann, W. 1977. Changes in bone and bone marrow of rabbits bearing the VX-2 carcinoma: A comparison of local and distal effects. *Am. J. Pathol.* 87:537–52
21. Seyberth, H. W., Hubbard, W. C., Oelz, O., Sweetman, B. J., Watson, J. T., Oates, J. A. 1977. Prostaglandin-mediated hypercalcemia in the VX_2 carcinoma-bearing rabbit. *Prostaglandins*. 14:319–31
22. Flower, R. J. 1974. Drugs which inhibit prostaglandin biosynthesis. *Pharmacol. Rev.* 26:33–67
23. Tashjian, A. H. Jr. 1975. Prostaglandins, hypercalcemia and cancer. *N. Engl. J. Med.* 293:1317–18
24. Raisz, L. G., Sandberg, A. L., Goodson, J. M., Simmons, H. A., Mergenhagen, S. E. 1974. Complement-dependent stimulation of prostaglandin synthesis and bone resorption. *Science* 185:789–91
25. Dowsett, M., Eastman, A. R., Easty, D. M., Easty, G. C., Powles, T. J., Neville, A. M., 1976. Prostaglandin mediation of collagenase-induced bone resorption. *Nature* 263:72–74
26. Ferreira, S. H., Vane, J. R. 1967. Prostaglandins: Their disappearance from and release into the circulation. *Nature* 216:868–73
27. Piper, P. J., Vane, J. R., Wyllie, J. H. 1970. Inactivation of prostaglandin by the lung. *Nature* 225:600–4
28. Hamberg, M., Samuelsson, B. 1971. On the metabolism of prostaglandin E_1 and E_2 in man. *J. Biol. Chem.* 25:6713–21
29. Hook, R., Gillis, C. N. 1975. The removal and metabolism of prostaglandin E_1 by rabbit lung. *Prostaglandins* 9:193–201
30. Beliel, O. M., Singer, F. R., Coburn, J. W. 1973 Prostaglandins: Effect on plasma calcium concentration. *Prostaglandins* 3:237–41
31. Robinson, C. J., Parsons, J. A. 1974. Lack of hypercalcaemia on infusing prostaglandin E_2 to rats, argues against its suggested role in the "non-parathyroid hypercalcaemic tumour" syndrome. *J. Endocrinol.* 64:14–15
32. Franklin, R. B., Tashjian, A. H. Jr. 1975. Intravenous infusion of prostaglandin E_2 raises plasma calcium concentration in the rat. *Endocrinology* 97:240–43
33. Seyberth, H. W., Hubbard, W. C., Morgan, J. L., Sweetman, B. J., Oates, J. A. 1976. Prostaglandin mediated hypercalcemia in the VX_2 carcinoma-bearing rabbit. *Clin. Res.* 24:370A
34. Tashjian, A. H. Jr., Voelkel, E. F., Levine, L. 1977. Effects of hydrocortisone on the hypercalcemia and plasma levels of 13,14-dihydro-15-keto-prostaglandin E_2 in mice bearing the $HSDM_1$ fibro-sarcoma. *Biochem. Biophys. Res. Commun.* 74:199–207
35. Tashjian, A. H. Jr., Voelkel, E. F., Levine, L. 1977. Plasma concentrations of 13,14-dihydro-15-keto-prostaglandin E_2 in rabbits bearing the VX_2 carcinoma: Effects of hydrocortisone and indomethacin. *Prostaglandins*. 14:309–17
36. Adkins, D., Martin, T. J. 1977. Rat osteogenic sarcoma cells. Effects of some prostaglandins, their metabolites and analogues on cyclic AMP production. *Prostaglandins* 13:861–71
37. Brereton, H. D., Haluschka, P. V., Alexander, R. W., Mason, D. M., Keiser, H. R., De Vita, V. T. Jr. 1974. Indomethacin-responsive hypercalcemia in a patient with renal-cell adenocarcinoma. *N. Engl. J. Med.* 291:83–85
38. Seyberth, H. W., Sergre, G. V., Hamet, P., Sweetman, B. J., Potts, J. T. Jr., Oates, J. A. 1976. Char-

acterization of the group of patients with the hypercalcemia of cancer who respond to treatment with prostaglandin synthesis inhibitors. *Trans. Assoc. Am. Phys*. 89:92–104

39. Robertson, R. P., Baylink, D. J., Metz, S. A., Cummings, K. B. 1976. Plasma prostaglandin E in patients with cancer with and without hypercalcemia. *J. Clin. Endocrinol. Metab*. 43:1330–35
40. Robertson, R. P., Baylink, D. J., Marini, J. J., Adkison, H. W. 1975. Elevated prostaglandins and suppressed parathyroid hormone associated with hypercalcemia and renal cell carcinoma. *J. Clin. Endocrinol. Metab*. 41:164–67
41. Ito, H., Sanada, T., Katayama, T., Shimazaki, J. 1975. Indomethacin-responsive hypercalcemia. *N. Engl. J. Med*. 293:558–59
42. Dindogru, A., Gailani, S., Henderson, E. S., Wallace, H. J. Jr., Fitzpatrick, J. 1975. Indomethacin in hypercalcaemia. *Lancet* 2:365
43. Blum, I. 1975. Indomethacin in hypercalcaemia. *Lancet* 1:866
44. Cummings, K. B., Wheelis, R. F., Robertson, R. P. 1975. Prostaglandin: Increased production by renal cell carcinoma. *Surg. Forum* 26:572–74
45. Demers, L. M., Allegra, J. C., Harvey, H. A., Lipton, A., Luderer, J. R., Mortel, R., Brenner, D. E. 1977. Plasma prostaglandins in hypercalcemic patients with neoplastic disease. *Cancer* 39:1559–62
46. Benvenisti, D., Goldberg, H. 1975. Prostaglandins, indomethacin and hypercalcemia in neoplastic disease. *N. Engl. J. Med*. 292:1189
47. Coombes, R. C., Neville, A. M., Bondy, P. K., Powles, T. J. 1976. Failure of indomethacin to reduce hydroxyproline excretion or hypercalcemia in patients with breast cancer. *Prostaglandins* 12:1027–35
48. Mundy, G. R., Raisz, L. G., Cooper, R. A., Schechter, G. P., Salmon, S. E. 1974. Evidence for the secretion of an osteoclast stimulating factor in myeloma. *N. Engl. J. Med*. 291:1041–46
49. Dowsett, M., Easty, G. C., Powles, T. J., Easty, D. M., Neville, A. M. 1976. Human breast tumour-induced osteolysis and prostaglandins. *Prostaglandins* 11:447–60

Ann. Rev. Med. 1978. 29:31–40
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CARDIOVASCULAR COMPLICATIONS OF ORAL CONTRACEPTIVES ◆7263

Norman M. Kaplan, M.D.
Department of Internal Medicine, University of Texas Southwestern Medical School, Dallas, Texas 75235

Oral contraceptives have accomplished their purpose but with unexpected and often serious side effects. This paper details the various cardiovascular complications of oral contraceptives (OC or "the pill"), and describes the possible mechanisms. Despite the appropriate emphasis on these problems, the use of the pill is defended. Guidelines for reduction of cardiovascular complications from its use are provided.

Since its introduction in 1960, the pill has become the most popular form of contraception, used by more than half of US married women, aged 15 to 29, who use any form of contraception (1). An estimated 8 million women in the US and 50 million worldwide were taking the pill in 1976 (2). This widespread use of the pill and other effective contraceptives has accomplished the societal goal of significantly reducing unplanned and unwanted births (3). But along with this major benefit, a number of problems have arisen; those involving the cardiovascular system are both the most common and the most serious.

THE TYPES OF CARDIOVASCULAR COMPLICATIONS

Hypertension, heart attacks, strokes, and blood clots are the most common cardiovascular complications of OC. Altogether, they are responsible for a threefold increased risk of death from cardiovascular disease among pill users, which amounts to an excess annual mortality of around 20 per 100,000 women aged 15 to 44 (4). This figure seems more impressive when compared to a total annual mortality of 70 per 100,000 from all causes and of 1 per 100,000 for all causes related to childbearing for women of comparable age.

These figures were derived from data from 21 different countries and an excellent correlation between mortality trends and prevalence of OC use was found (4). Moreover, the risks seem to be just as large with the newer combination pills with lower doses of estrogen. The use of low-dose estrogen pills reduces, but does not eliminate, the risk of thromboembolic disease (5); it does not seem to affect the incidence of hypertension (6). The derangements respon-

0066-4219/78/0401-0031$01.00

sible for cardiovascular disease may develop with an amount of estrogen well below the contraceptive dose. Moreover, the dose of progestogen may also be implicated (6), so that a reduction in estrogen dose may not improve overall safety.

Hypertension

Of 23,000 women on the pill for five years, 5% developed hypertension, a rate 2.6 times greater than that seen among a comparable group of nonusers(6). All women who use estrogen-containing pills have a slight but significant rise in blood pressure, averaging 6/2 mm Hg in both cross-sectional and longitudinal studies of over 13,000 women in California (7). The rise in pressure may be progressive, as shown in the more carefully controlled study of a small group of women in Scotland (8) (Figure 1).

RELATION TO TYPE OF CONTRACEPTIVE The culprit is likely the estrogen since, in most studies with progestogens alone, neither biochemical abnormalities nor rises in blood pressure have been seen (9). However, the Royal College study (6) found no relation between the incidence of hypertension and the dose of estrogen, which varied between 50–150 μg per day, but did find a relation with the dose of progestogen. Women taking pills with only 20–30 μg of estrogen were not included in this study.

PREDISPOSING FACTORS The known predisposing factors identified in various studies are (*a*) increasing age, (*b*) obesity, and (*c*) a positive family history of hypertension. Some find an association with prior preeclampsia to be a predisposing factor (10), others do not (11). Mexican-American women are more susceptible, perhaps related to their greater degree of obesity (12). Though it seems prudent not to give the pill to women already hypertensive, no further rise in blood pressure was observed in 33 hypertensive women given OC for one year (13).

NATURE OF THE HYPERTENSION In most women, the rise in blood pressure is minimal but, in some, severe and rapidly accelerating hypertension may cause irreversible renal damage (14, 15). Even with reversible hypertension, considerable renal damage may persist (16).

In about half of the users, blood pressure returns to normal when the pill is stopped. Whether the pill causes new hypertension or simply uncovers essential hypertension that would eventually appear spontaneously is unknown. Among 14 women whose blood pressure became normal after the pill was stopped, 7 developed spontaneous hypertension during the subsequent 6 years (17).

MECHANISMS AND CAUSES OF HYPERTENSION About 100–160 mEq of sodium are retained (18) and blood volume and cardiac output increase, even in women whose blood pressure does not increase (19, 20).

Estrogen increases hepatic synthesis of the protein substrate upon which the enzyme renin acts to produce the pressor hormone, angiotensin (21). The presence of increased substrate results in increased levels of plasma renin

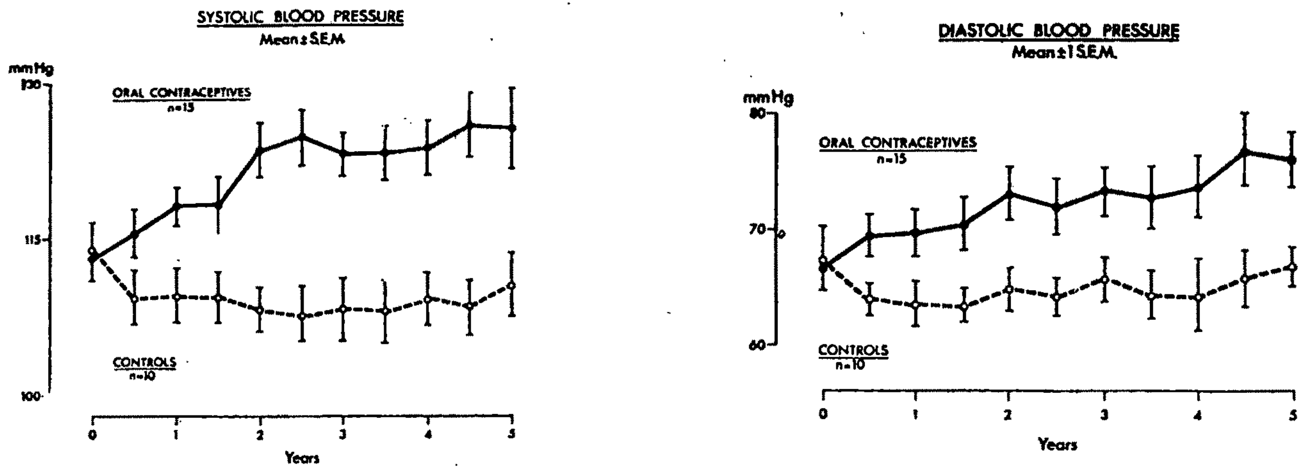


Figure 1 Changes in systolic (*right*) and diastolic (*left*) blood pressure after 5 yr in women taking oral contraceptives and in a control group of women using intrauterine contraceptive devices or cervical diaphragms (8).

activity (22) and plasma angiotensin II (23). These increased levels of renin-angiotensin stimulate adrenal synthesis of aldosterone, which likely is responsible for the sodium retention.

The logical assumption has been made that the hypertension reflects the activated renin-angiotension-aldosterone mechanism. However, with few exceptions (24), identical changes are seen in those who do not develop hypertension. It may be that women vary in their vascular and renal responsiveness to these changes, those with greatest responsiveness developing hypertension. Such variability has been measured in the degree of angiotensin-induced renal vasoconstriction (25). A role for angiotensin is supported by evidence that the elevated blood pressure of estrogen-treated rats (26) is lowered by the angiotensin antagonist, saralasin. Similar effects have been noted in women made hypertensive by oral contraceptives (27).

Other mechanisms may be involved. The levels of plasma dopamine-β-hydroxylase activity, an index of sympathetic nervous system activity, were increased in 25 women who developed hypertension on the pill (28).

Whatever the mechanisms, estrogen-induced hypertension is likely to be a factor in the various vascular catastrophes seen with use of the pill. As in other forms of hypertension, the elevated blood pressure aggravates and accelerates the atherosclerotic process and may cause erosion and rupture of vessels in a direct manner, related to the level of pressure (29).

Thromboembolism

An increased tendency for the blood to clot may also be an important mechanism by which the pill induces vascular disease, both directly within leg veins and various arteries, and indirectly by obstructing blood flow to vital organs by embolization and thrombus formation.

RISK OF THROMBOSIS AND EMBOLISM Most researchers have found an increased risk of both peripheral venous thrombosis and pulmonary embolism in OC users, the relative increase varying from two- to twelvefold. Various epidemiological techniques have been used to identify this increased risk. These include:

1. retrospective case-control studies such as those from the Boston Collaborative Drug Surveillance Program (30), and from Sartwell and coworkers (31) showing an eleven- and fourfold increased risk, respectively;
2. review of trends in mortality statistics, which has shown an increase in mortality in OC users (32);
3. reports of adverse reactions, which likely underestimate the incidence, but also show a definite increase in risk (33); and
4. prospective trials, the largest of which, currently under way in England, has found a 5.6 times increased risk of deep leg vein thrombosis (6).

The advent of lower-dose estrogen pills seems to have decreased the risk by at least 25% (6, 32), but women taking only 30–50 μg estrogen pills continue to have more deep vein thrombosis (34).

Some people question the relation between OC use and thromboembolism; they point out the inability of retrospective case-control studies and reports of adverse reactions to establish a cause-and-effect relationship, and the possibility that other predisposing risk factors may be more common in pill users who develop thromboembolism (35). However, except for one study (36) that failed to find an increased incidence, the data overwhelmingly support a greater risk for thromboembolism in women who take the pill than for those who do not.

MECHANISMS FOR INCREASED THROMBOGENICITY A number of possible mechanisms by which estrogens increase thrombus formation have been described. These include:

1. enhancement of coagulation by decreasing the negative surface charge on vessel walls and blood cells (37);
2. a decrease in the linear velocity of blood flow induced by the greater distensibility of peripheral veins (38); and
3. changes in the coagulation and fibrinolytic systems which favor clotting, including decreased plasma antithrombin III activity (34) and reduced activated factor X inhibitory activity (39).

These changes are presumably responsible for various abnormalities in laboratory tests: shorter clotting times and firmer clots as demonstrated by thromboelastography (40); shorter activated thromboplastin times (41); chromatographically demonstrable plasma fibrinogen complexes (42); and a low content of fibrinolytic activators in venous vessel walls (43).

In addition to these changes which could favor clotting, arteries may be thickened and roughened by structural changes induced by intimal proliferation or deposition of atheroma. Structural changes of endothelial proliferation and intimal thickening were first reported in women who died suddenly from a thromboembolic event while taking OC (44). More recently, similar changes were found in arteries from uteri removed from living women (45). Atheroma deposition may be favored by increased serum triglycerides and decreased glucose tolerance. Serum triglyceride concentration usually increases with OC use, occasionally to very high levels (46). The mechanism involves increased triglyceride entry into the plasma (47), perhaps by estrogen induction of hepatic microsomal enzymes involved in triglyceride synthesis (48). Glucose tolerance also usually worsens with OC use (49), but rarely to the point of fasting hyperglycemia or overt diabetes. The long-term contributions of both hypertriglyceridemia and decreased glucose tolerance to the vascular diseases seen with OC use remain unknown.

Coronary Artery Disease

Hypertension, increased thrombogenicity, and structural damage to the vessel walls could accelerate coronary artery disease. English women on the pill have more fatal and nonfatal myocardial infarcts than comparably aged women who do not use the pill (50–52); the increase in risk is about threefold greater. The risk from OC use acts synergistically with cigarette smoking, hypertension, diabetes, and hypercholesterolemia; women with 3 or more of these risk factors have a 128-fold increased risk of myocardial infarction (53).

Analysis of the above data and other reports on the association between OC use and myocardial infarction led to these conclusions (54): In women over age 30, the use of the pill in the absence of other predisposing factors increases the risk of dying from myocardial infarction about twofold, of the same order of magnitude as the increased risk of death from thromboembolism; women 30 to 39 years of age who have any one of the three main predisposing factors (smoking, hypertension, hypercholesterolemia) have a threefold increased risk and should be warned against using OC; women over age 40 should not take the pill.

The Boston Collaborative Drug Surveillance Program found a 1.9 relative risk for myocardial infarction among premenopausal OC users (55). When various additional factors were considered, the risk ratio for OC use alone decreased to only 1.3.

Beyond the various mechanisms previously considered, estrogens may increase coronary artery smooth muscle tone, causing myocardial ischemia by coronary spasm (56). Of 53 patients previously found to have ST-segment abnormalities, 45 had greater ST depression after 2 weeks' intake of estrogen. The pretreatment appearances were regained 6 weeks after the estrogen intake was stopped.

Cerebral Vascular Disease

The results of a collaborative study (57) of 598 women, aged 15 to 44, who had had a stroke, indicate a 9.5 times greater risk of thrombotic stroke and a 2.0 times greater risk of hemorrhagic stroke among pill users. OC use alone was found to be a significant risk for stroke, in the absence of smoking, hypertension, or migraine (58). The combination of severe hypertension (greater than 180/110 mm Hg) and OC use markedly increased the risk of hemorrhagic stroke. Vertebral artery occlusion is also increased with pill use (59), and migraine may worsen as well (60). However, despite these increased risk ratios, the absolute risk is relatively low, estimated to be 10 per 100,000 per year among women using the pill.

Other Cardiovascular Complications

Pulmonary hypertension of unknown cause was reported in six women using the pill (61). Peripheral arterial occlusion (62) and mesenteric vascular insufficiency with massive colonic bleeding (63) have also been seen.

CONCLUSIONS CONCERNING CARDIOVASCULAR COMPLICATIONS

The pill is safe for most young women for temporary birth control and family planning. It should not be used in women over age 35 who want no more children. Significant problems (such as hypertension, increased tendency to clot, rise in triglyceride, and decrease in glucose tolerance) will occur in 5–10% of women of any age who take the pill for 5 years. These factors are, in turn, responsible for various cardiovascular complications (such as thromboembolism, heart attack, and stroke) which occur with increased frequency among pill users. The combination of all these gives more excess mortality than the number of lives saved by prevention of unwanted pregnancy (52).

Most women who take the pill will suffer more subtle disturbances in blood pressure, biochemical, and hormonal levels, which may pose additional risks for subsequent cardiovascular disease. Much longer surveillance of larger numbers of women will be needed to settle the issues.

Recommendations for Oral Contraceptive Use

The following recommendations seem consonant with current information about the safety of the pill:

1. Recommend alternate forms of contraception to those who are above age 35, obese, hypertensive, or smoke cigarettes.
2. If the pill is chosen for such women, encourage weight reduction, control of hypertension, and cessation of smoking.
3. Never provide more than a six-month supply without requiring that the woman return for surveillance.
4. Always check the blood pressure on return visits. If it goes above 140/90 mm Hg, stop the pill if possible. If it cannot be stopped, keep the blood pressure down with appropriate antihypertensive therapy, starting with a diuretic.
5. If there is a family history of cardiovascular disease, test serum triglyceride and cholesterol levels before and after three to six months use of the pill. If they are elevated, stop the pill.
6. Consider use of low-dose heparin therapy whenever a women using OC undergoes trauma, surgery, or prolonged bed rest.

In summary, the risks of pill use often outweigh its benefits. Prudent use of oral contraceptives will improve the balance between risk and benefit.

Literature Cited

1. Ford, K. 1976. Contraceptive utilization among currently married women 15–44 years of age: United States, 1973. *Monthly Vital Statistics Report* 25 (No. 7): Suppl. pp. 1–24
2. Rinehart, W., Ravenhold, R. T. 1977. U.S. morbidity and mortality trends relative to oral contraceptive use, 1955–1975, and Danish morbidity trends, 1953–1972. *Population Reports:* Suppl. to Series A, No. 4. pp. 105–32
3. Westoff, C. F. 1976. The decline of unplanned births in the United States. *Science* 191:38–41
4. Beral, V. 1976. Cardiovascular-disease mortality trends and oral-contraceptive use in young women. *Lancet* 2:1047–51
5. Stolley, P. D., Tonascia, J. A., Tockman, M. S., Sartwell, P. E., Rutledge, A. H., Jacobs, M. P. 1975. Thrombosis with low-estrogen oral contraceptives. *Am. J. Epidemiol.* 102:197–208
6. *Oral Contraceptives and Health,* an interim report from the Oral Contraceptive Study of the Royal College of General Practitioners. 1974. pp. 37–42. New York: Pitman
7. Fisch, I. R., Frank, J. 1977. Oral

contraceptives and blood pressure. *J. Am. Med. Assoc.* 237:2499–2503

8. Weir, R. J., Davies, D. L., Fraser, R., Morton, J. J., Tree, M., Wilson, A. 1975. Contraceptive steroids and hypertension. *J. Steroid Biochem.* 6:961–64
9. Spellacy, W. N., Birk, S. A. 1972. The effect of intrauterine devices, oral contraceptives, estrogens, and progestogens on blood pressure. *Gynecology* 112:912–19
10. Mason, B., Oakley, N., Wynn, V. 1973. Studies of carbohydrate and lipid metabolism in women developing hypertension on oral contraceptives. *Br. Med. J.* 3:317–20
11. Pritchard, J. A., Crosby, U. D., Martin, F. G., Pritchard, S. A. 1974. Blood pressure during estrogen + progestin contraception after pregnancy-induced hypertension. *Gynecol. Invest.* 5:26
12. Stern, M. P., Brown, B. W. Jr., Haskell, W. L., Farquhar, J. W., Wehrle, C. L., Wood, P. D. S. 1976. Cardiovascular risk and use of estrogens or estrogen-progestogen combination. *J. Am. Med. Assoc.* 235:811–15
13. Spellacy, W. N., Birk, S. A. 1974. The effects of mechanical and steroid contraceptive methods on blood pressure in hypertensive women. *Fertil. Steril.* 25:467–70
14. Zacherle, B. J., Richardson, J. A. 1972. Irreversible renal failure secondary to hypertension induced by oral contraceptives. *Ann. Intern. Med.* 77:83–85
15. Schoolwerth, A. C., Sandler, R. S., Klahr, S., Kissane, J. M. 1976. Nephrosclerosis postpartum in women taking oral contraceptives. *Arch. Intern. Med.* 136:178–85
16. Boyd, W. N., Burden, R. P., Aber, G. M. 1975. Intrarenal vascular changes in patients receiving oestrogen-containing compounds—A clinical, histological and angiographic study. *Q. J. Med.* 44:415–31
17. Woods, J. W., Algary, W. A., Stier, F. M. 1972. Oral contraceptives and hypertension. *Circulation* 45–46: Suppl. II, p. 82 (Abstr.)
18. Crane, M. G., Harris, J. J. 1973. Effect of estrogens and gestagens on the renin-aldosterone system. In *Oral Contraceptives and High Blood Pressure*, ed. M. J. Fregly, pp. 100–19. Gainesville: Dolphin
19. Walters, W. A. W. 1970. Haemodynamic changes in women taking oral contraceptives. *J. Obstet. Gynaecol. Br. Commonw.* 77:1007–12
20. Lehtovirta, P. 1974. Haemodynamic effects of combined oestrogen/progestogen oral contraceptives. *J. Obstet. Gynaecol. Br. Commonw.* 81:517–25
21. Helmer, O. M., Griffith, R. S. 1952. The effect of the administration of estrogens on the renin-substrate (hypertensinogen) content of rat plasma. *Endocrinology* 57:421–26
22. Skinner, S. L., Lumbers, E. R., Symonds, E. M. 1969. Alteration by oral contraceptives of normal menstrual changes in plasma renin activity, concentration and substrate. *Clin. Sci.* 36:67–76
23. Cain, M. D., Walters, W. A., Catt, K. J. 1971. Effects of oral contraceptive therapy on the renin-angiotensin system. *J. Clin. Endocrinol.* 33:671–76
24. Saruta, T., Saade, G. A., Kaplan, N. M. 1970. A possible mechanism for hypertension induced by oral contraceptives. *Arch. Intern. Med.* 127:621–26
25. Hollenberg, N. K., Williams, G. H., Burger, B., Chenitz, W., Hoosmand, I., Adams, D. F. 1976. Renal blood flow and its response to angio-tensin II. *Circ. Res.* 38:35–40
26. Stubbs, D. H., Johnson, J. A., Keitzer, M. F. 1976. Effect of 1-sarcosine-8-alanine angiotensin II on arterial pressure in oral contraceptive-induced hypertension in rats. *Fed. Proc.* 35:398 (Abstr.)
27. Streeten, D. H. P., Anderson, G. H. Jr., Dalakos, T. G. 1976. Angiotensin blockade: Its clinical significance. *Am. J. Med.* 60:817–24
28. Rockson, S. G., Stone, R. A., Gunnells, J. C., Schanberg, S. M., Kirshner, N., Robinson, R. R. 1975. Plasma dopamine-ß-hydroxylase activity in oral contraceptive hypertension. *Circulation* 51:916–23
29. Kaplan, N. M. 1977. *Clinical Hypertension*, pp. 305–26. Baltimore: Williams & Wilkins
30. Boston Collaborative Drug Sur-

veillance Program. 1973. Oral contraceptives and venous thromboembolic disease, surgically confirmed gallbladder disease, and breast tumors. *Lancet* 1:1399–1404

31. Sartwell, P. E. 1971. Oral contraceptives and thromboembolism: A further report. *Am. J. Epidemiol.* 94:192–201
32. Sartwell, P. E., Stolley, P. D., Tonascia, J. A., Tockman, M. S., Rutledge, A. H., Wertheimer, D. 1976. *Prev. Med.* 5:15–19
33. Bottinger, L. E., Westerholm, B. 1971. Oral contraceptives and thromboembolic disease. *Acta Med. Scand.* 190:455–63
34. Sager, S., Thomas, D. P., Stamatakis, J. D., Kakkar, V. V. 1976. Oral contraceptives, antithrombin-II activity, and postoperative deep-vein thrombosis. *Lancet* 1:509–11
35. Kay, C. R., Smith, A., Richards, B. 1969. Smoking habits of oral contraceptive users. *Lancet* 2:1228–29
36. Fuertes-de la Haba, A., Curet, J. O., Pelegrina, I. 1971. Thrombophlebitis among oral and nonoral contraceptive users. *Obstet. Gynecol.* 38:259–63
37. Srinivasan, S., Solash, J., Redner, A., Moser, C., Farhangian, D., Lucas, T. R., Ateyeh, E. Jr., McCool, C. T., Weber, D., Puvan, S. I., Jones, J. R., Sawyer, P. N. 1974. The alteration of surface charge characteristics of the vascular system by oral contraceptive steroids. *Contraception* 9:291–303
38. Goodrich, S. M., Wood, J. E. 1964. Peripheral venous distensibility and velocity of venous blood flow during pregnancy or during oral contraceptive therapy. *Am. J. Obstet. Gynecol.* 90:740–44
39. Wessler, S., Gitel, S. N., Wan, L. S., Pasternack, B. S. 1976. Estrogen-containing oral contraceptive agents. A basis for their thrombogenicity. *J. Am. Med. Assoc.* 236:2179–82
40. Fisch, I. R., Freedman, S. H., Pellegrin, F. A. 1973. Effect of oral contraceptives on the thromboelastogram. *Clin. Pharmacol. Ther.* 14:238–44
41. Howie, P. W., Prentice, C. R. M., Mallinson, A. C., Horne, C. H. W., McNicol, T. P. 1970. Effect of combined oestrogen-progestogen oral contraceptives, oestrogen, and progestogen on antiplasmin and antithrombin activity. *Lancet* 2:1329–32
42. Alkjaersig, N., Fletcher, A., Burstein, R. 1975. Association between oral contraceptive use and thromboembolism: A new approach to its investigation based on plasma fibrinogen chromatography. *Am. J. Obstet. Gynecol.* 122:199–211
43. Astedt, B. 1975. New aspects of the thrombogenic effect of oral contraceptives. *Am. Heart J.* 90:1–3
44. Irey, N. S., Manion, W. C., Taylor, H. B. 1970. Vascular lesions in women taking oral contraceptives. *Arch. Pathol.* 89:1–8
45. Osterholzer, H. O., Grillo, D., Kruger, P. S., Dunnihoo, D. R. 1977. The effect of oral contraceptive steroids on branches of the uterine artery. *Obstet. Gynecol.* 49:227–32
46. Davidoff, F., Tishler, S., Rosoff, C. 1973. Marked hyperlipidemia and pancreatitis associated with oral contraceptive therapy. *N. Engl. J. Med.* 289:552–55
47. Kim, H., Kalkhoff, R. K. 1975. Sex steroid influence on triglyceride metabolism. *J. Clin. Invest.* 56:888–96
48. Martin, J. V., Martin, P. J., Goldberg, D. M. 1976. Enzyme induction as a possible cause of increased serum-triglycerides after oral contraceptives. *Lancet* 1:1107–8
49. Posner, N. A., Silverstone, F. A., Tobin, E. H. 1975. Changes in carbohydrate tolerance during long-term oral contraception. *Am. J. Obstet. Gynecol.* 123:119–27
50. Mann, J. I., Vessey, M. P., Thorogood, M., Doll, R. 1975. Myocardial infarction in young women with special reference to oral contraceptive practice. *Br. Med. J.* 2:241–45
51. Mann, J. L., Thorogood, M., Waters, W. E., Powell, C. 1975. Oral contraceptives and myocardial infarction in young woman: A further report. *Br. Med. J.* 3:631–32
52. Mann, J. I., Inman, W. H. W., Thorogood, M. 1976. Oral contraceptive use in older women and fatal myocardial infarction. *Br. Med. J.* 2:445–47
53. Mann, J. I., Doll, R., Thorogood,

M., Vessey, M. P., Waters, W. E. 1976. Risk factors for myocardial infarction in young women. *Brit. J. Prev. Soc. Med.* 30:94–100

54. Ory, H. W. 1977. Association between oral contraceptives and myocardial infarction. *J. Am. Med. Assoc.* 237:2619–22
55. Rosenberg, L., Armstrong, B., Jick, H. 1976. Myocardial infarction and estrogen therapy in premenopausal women. *N. Engl. J. Med.* 294:1290–91
56. Jaffe, M. D. 1977. Effect of oestrogens on postexercise electrocardiogram. *Br. Heart J.* 38:1299–1303
57. Collaborative Group for the Study of Stroke in Young Women. 1973. Oral contraception and increased risk of cerebral ischemia or thrombosis. *N. Engl. J. Med.* 288:871–78
58. Collaborative Group for the Study of Stroke in Young Women. 1975. Oral contraceptives and stroke in young women. *J. Am. Med. Assoc.* 231:718–22
59. Ask-Upmark, E., Bickerstaff, E. R. 1976. Vertebral artery occlusion and oral contraceptives. *Br. Med. J.* 1:487–88
60. Whitty, C. W. M., Hockaday, J. M., Whitty, M. M. 1966. The effect of oral contraceptives on migraine. *Lancet* 1:856–59
61. Kleiger, R. E., Boxer, M., Ingham, R. E., Harrison, D. C. 1976. Pulmonary hypertension in patients using oral contraceptives: A report of six cases. *Chest* 69:143–47
62. O'Connor, J., Pollock, J. G. 1975. Acute arterial occlusion and the contraceptive pill. *J. Cardiovas, C. Surg.* 16:176–80
63. Prust, F. W., Kumar, G. K. 1976. Massive colonic bleeding and oral contraceptive "pills." *Am. J. Obstet. Gynecol.* 125:695–98

Ann. Rev. Med. 1978. 29:41–49
Copyright © 1978 by Annual Reviews Inc. All rights reserved

RELATIONSHIPS BETWEEN PLATELETS AND COAGULATION FACTORS IN HEMOSTASIS[1]

◆7264

Philip W. Majerus, M.D. and Joseph P. Miletich
Division of Hematology-Oncology, Departments of Internal Medicine and
Biochemistry, Washington University School of Medicine, St. Louis, Missouri 63110

INTRODUCTION

In this review the relations between platelets and coagulation factors in hemostasis are considered. The traditional view of hemostasis implies that platelets aggregate to form a hemostatic plug in wounds prior to the coagulation reactions which lead to fibrin formation. In this view the platelet plug merely provides a phospholipid surface on which the coagulation reactions take place. In contrast, we present findings that suggest much closer interactions between platelets and coagulation factors. For example, thrombin is formed very early in hemostasis; thus a coagulation factor may be involved in early reactions of platelet aggregation. Also, activation of coagulation factors may take place on the platelet surface on specific sites or receptors that direct the coagulation reactions; that is, platelets may provide more than a simple phospholipid surface to the coagulation sequence.

ROLE OF PLATELETS IN HEMOSTASIS

The first step in hemostasis, which can be defined morphologically when a vessel is cut and a wound is produced, is the adhesion of platelets to collagen or other substances found in the tissues beneath the vascular endothelium. This interaction does not explain the major growth of the platelet plug, however, since most platelets do not come into contact with the wound surface. Subsequent to adhesion to the wound surface, the platelets produce arachidonic acid metabolites, particularly prostaglandin G_2 and thromboxane A_2 (1). These labile substances, and ADP from platelet storage granules, stimulate aggregation of other

---

[1] This research was supported by grants HL 14147 (Specialized Center for Research in Thrombosis) and HL 16634 from the National Institutes of Health.

0066-4219/78/0401-0041$01.00

platelets and a further release reaction. The enzymes that form prostaglandin G_2 and thromboxane A_2 utilize arachidonic acid cleaved from platelet phospholipids by a phospholipase A_2 (2). Agents that stimulate the platelet release reaction also activate this phospholipase. Aspirin covalently acetylates and inactivates the first enzyme in the prostaglandin biosynthetic pathway, which is termed "arachidonic acid cyclo-oxygenase" (3, 4). This action of aspirin blocks formation of the active arachidonate metabolites and explains its antiplatelet effect (5, 6). Thromboxane A_2 produced by platelets is a potent vasoconstrictor (greater than angiotensin II) and therefore constricts the cut vessel. Other vasoactive substances, including serotonin, are secreted from platelet granules [for recent review, see (7)]. Coagulation factors are also released from granules contained within the platelets; these include factor V and fibrinogen as well as a heparin-neutralizing substance called platelet factor 4. The precise role of these factors in the formation of the hemostatic plug is somewhat obscure and, in fact, it seems paradoxical that platelets should secrete substances that are contained in high concentrations in the surrounding plasma, i.e. fibrinogen, factor V, etc. Possibly these factors are released from platelets to achieve high local concentrations in the platelet aggregate. It is also possible that the released forms of these molecules are different from their plasma counterparts, although the characterization in the case of fibrinogen suggests that platelet fibrinogen is identical to plasma fibrinogen (8). Another substance released by the platelet is a heat-stable protein that is required for the proliferation of smooth-muscle cells and fibroblasts in tissue culture (9, 10). This substance may function in vivo as the stimulus to initiate wound healing.

According to most workers, the platelet plug provides the initial hemostatic vehicle, and only subsequently do coagulation reactions leading to fibrin formation occur on the platelet phospholipid surface. An alternative hypothesis suggests that the separation of platelet reactions from blood coagulation is artificial and that the two are intimately related. In fact, it has been proposed that a minute amount of thrombin, which is formed prior to generation of a fibrin clot (11, 12), is the stimulus for the platelet release reaction and the production of prostaglandin metabolites. A number of potentially important hemostatic reactions are catalyzed by concentrations of thrombin well below that required to clot fibrinogen. These include the activation of factors V and VIII as well as the initiation of platelet aggregation and the release reaction. Furthermore, recent experimental findings (see below) suggest that thrombin formation may take place on the platelet surface so that the local concentration of thrombin available to stimulate platelet secretion and aggregation may be very high in the area of the forming hemostatic plug.

INTERACTION OF THROMBIN WITH PLATELETS

Support for the concept that coagulation factors interact with platelets early in hemostasis came from recent experiments, which utilized a radioimmunoassay for thrombin. This method allowed measurements of small amounts of thrombin in whole blood and showed that thrombin is formed within a few seconds of venipuncture. Low concentrations of thrombin in the range of 3–10 ng/ml were present for many minutes prior to the clotting of blood in a test tube (12). In similar experiments the platelet release reaction was monitored in clotting whole blood by radioimmunoassay of platelet factor 4, and these experiments demonstrated that platelet release also occurs only a few seconds after venipuncture in parallel with thrombin generation (M. A. Shuman and S. Levine, unpublished observations). However, these experiments can only be a very distant model of what must happen in a wound in vivo where clotting occurs much more rapidly than in the test tube, and where it is clear from a variety of morphometric studies that platelet adhesion and aggregation is the first event in the hemostatic process. Additionally, almost none of the reactions studied in vitro have taken into account the many coagulation inhibitors that are present in plasma, although these substances are extremely important in limiting or regulating the hemostatic process. Thus, when a vessel is wounded, a platelet plug and a fibrin clot form in the tissues outside the vessel wall but do not extend into the lumen to occlude the blood vessel itself. This indicates that the growth of the hemostatic plug and fibrin clot must be closely regulated. The mechanism of this fine control is completely unknown. However, a number of recent experiments do indicate close relationships between reactions of platelets and various clotting factors, and these are considered in some detail below.

The recent availability of homogeneous preparations of human clotting factors has allowed for study of the direct interactions between thrombin and platelets (13–15). These experiments were carried out by radioiodinating the pure thrombin, which thereby allows measurements of direct binding to platelets. In this way, we have shown that thrombin binds to the surface of intact human platelets at specific receptor sites (13), and that the binding appears to be the first reaction in the process of the thrombin-stimulated platelet aggregation and release reactions (16). Electron microscopy and autoradiography show (13) that thrombin binds to the cell surface in the outer glycoprotein coat of the platelet.

Thrombin binding does not demonstrate typical saturation kinetics. As increasing numbers of thrombin molecules bind to the cells, the affinity or ability of subsequent molecules to bind decreases. This type of binding has been termed "negative cooperativity" and suggests at a physiological level that there may be a threshold-type response of platelets to thrombin, i.e. all of the platelet receptors are saturated to a certain level before any response is obtained (17, 18). This mechanism, if applicable in vivo, might tend to dampen or prevent unwanted thrombin-platelet interactions. Negative-cooperative binding has also been observed in studies of several hormones and their associated receptor sites, although the physiological significance in those cases is obscure (19). The number of thrombin-binding sites per platelet is about 30,000–40,000, although only about 500 of these are the very high affinity type that seem to be associated with platelet aggregation and the release reaction. Even within these few functionally important sites, marked negative cooperativity of binding is observed. When thrombin-binding studies are carried out in parallel with measurements of the platelet release reaction, it is clear that there is very close correlation between the amount of thrombin bound under various experimental conditions and the extent of the platelet release reaction. When 100–200 molecules of thrombin have been bound to the platelet surface, the release reaction is essentially complete (16). Thrombin inactivated with diisopropyl fluorophosphate (DIP thrombin) binds to platelets with the same affinity and to the same extent as native thrombin, but this active site–inhibited thrombin is unable to stimulate the platelet secretion reaction (13). This observation has led a number of investigators (20–22) to suggest that thrombin stimulates platelet secretion by cleaving some protein on the platelet surface; however, no thrombin substrate has ever been directly demonstrated, and the kinetics of the thrombin reaction suggest that proteolysis might not be required. It is possible that DIP thrombin binds to the platelet surface but is unable to stimulate secretion because of the bulky diisopropyl substituent, not because of the loss of the proteolytic catalytic site. Only further experiments can settle this point. The thrombin receptor site on the platelet surface is highly specific. Other proteolytic enzymes, such as trypsin, that also stimulate platelet aggregation and release do so by a mechanism distinct from that of thrombin. Ten- to fiftyfold more trypsin than thrombin is required for the reaction, and binding studies using trypsin indicate no competition for thrombin-binding sites (17).

Prothrombin does not compete with thrombin for binding. The various activation intermediates involved in the conversion of prothrombin to thrombin, including prethrombin II, also do not compete. Prethrombin II has the same amino acid sequence as thrombin but lacks a single internal cleavage required to convert the single-chain prethrombin II to the active two-chain thrombin molecule (23). This indicates that the specificity of the thrombin-binding site requires the native thrombin molecule. Further support for this concept comes from studies that used various proteolyzed thrombins, none of which approach native thrombin in ability to stimulate human platelets (24). Thrombin from rabbits and cows also binds very tightly to human platelets. The platelet-binding regions of these three species seem highly conserved, which suggests that the platelet receptor site is physiologically important (25). That thrombin is actually formed on the platelet surface is suggested below.

ROLE OF COAGULATION FACTORS

Over the past 10–20 years the plasma coagulation factors have been identified and characterized. Most of these substances have recently been purified to homogeneity (see 26). The sequence of coagulation that is most widely accepted is called the waterfall or cascade hypothesis (27, 28). According to this scheme, coagulation is divided into intrinsic and extrinsic systems. The roles of these two systems in relation to each other is not known, although some investigators believe that the extrinsic system of factor VII and tissue factor is the initial stimulus that sets off the coagulation system. The intrinsic system serves as a positive feedback system that allows for the generation of the larger amounts of thrombin required to clot fibrinogen (29). The sequence of reactions in the intrinsic system is somewhat arbitrary and is based almost entirely on in vitro experiments. For example, when plasma is recalcified or blood is allowed to clot in glass or plastic tubes, factor XII is the earliest participant in the coagulation reaction, and the contact activation of factor XII leads to the sequential activation of factors XI, IX, and X, prior to thrombin formation. However, in vivo it is not clear that this same sequence obtains. Indeed, patients with factor XII deficiency have no hemostatic abnormality. Furthermore, according to the cascade hypothesis stated above, platelets and factors V and VIII (30, 31) are activated by thrombin in reactions apparently proximal to thrombin formation. Some investigators postulate that when a wound occurs, tissue factor is exposed to factor VII and these two substances activate a small amount of factor X, which in turn leads to formation of small amounts of thrombin. This thrombin activates factors V and VIII to initiate the intrinsic system of coagulation, which generates the large additional amounts of thrombin necessary to produce a fibrin clot (29).

An alternative hypothesis, most clearly elucidated by Walsh (32), suggests that the primary events of blood clotting occur on the platelet surface rather than in solution or on some other surface in a wound. This hypothesis is not suitably analyzed by traditional methods of studying coagulation factors in which various artificial mixtures of phospholipids are substituted for platelets.

FORMATION OF THROMBIN ON THE PLATELET SURFACE

A number of experiments have indicated that phospholipids and factor V can replace the need for platelets in prothrombin activation in vitro (33). Additionally, phospholipid can substitute for platelets in various assays that measure the clotting time of activated plasma. For example, addition of Russell's viper venom activates factor X, and the clotting of such activated plasma can be shortened by addition of either platelets or phospholipids (34, 35). Thus, it has been proposed that the entity in the platelet that promotes the conversion of prothrombin to thrombin is a phospholipoprotein on the platelet surface, which historically has been termed platelet factor 3. A variety of empirical tests have been devised to measure this substance, but because its exact chemical nature and subcellular localization have never been clearly defined, the term should probably not be used. The recent availability of homogeneous preparations of human factor X and X_a has allowed us to study the possible interaction of these clotting factors with platelets and the role of the platelets in prothrombin activation (36). In our preliminary experiments we showed that neither factor X nor its activated derivative, factor X_a, bound to washed human platelets in experiments similar to those that demonstrated the binding of thrombin to platelets. Also, factor X_a did not stimulate the release reaction. These experiments were carried out in widely varying concentrations of radioiodinated factors X and X_a and under a variety of conditions, including changes in the concentration of calcium ions. In contrast, both factor X and factor X_a bind to negatively charged phospholipid micelles in the presence of calcium ions [for review see (37)]. When prothrombin was included in mixtures of platelets with calcium ions and radioactive factor X_a, there was a slow formation of small amounts of thrombin as would be expected (slow, because no factor V was added). Under these conditions, the platelets released serotonin because of the small amount of thrombin that was bound. Factor X_a began to bind to the platelets only after the release reaction. It thus appears that receptor sites are exposed on the platelet surface only after the release reaction. When platelets were treated with a small amount of thrombin (0.5 U/ml) for a few minutes prior to the addition of factor X_a to stimulate the release reaction, binding of factor X_a to the platelet surface occurred immediately and reached equilibrium within a few minutes. The analysis of this binding shows that factor X_a binds with very high affinity to the platelet surface, and that approximately 200 factor X_a–binding sites per platelet are exposed.

Thrombin is not required for generation of X_a receptor sites, since the release reaction can be stimulated by either arachidonic acid or the calcium ionophore A23187. In both cases X_a-binding sites appeared. Once factor X_a was bound to the platelet surface, the rate of thrombin generation accelerated markedly. Factor X_a bound to the platelet surface formed thrombin 60,000-fold faster than factor X_a in solution. An optimal amount of phospholipid and factor V added in place of platelets resulted in rates of thrombin generation approximately 10- to 100-fold lower than those obtained by using platelets (37). Phospholipid alone was 1000-fold less effective than platelets. Thus, it is clear that the platelet provides more than a simple phospholipid surface in catalyzing the formation of thrombin. This is also shown in binding experiments where an excess of the unlabeled factor X zymogen does not compete with factor X_a for binding to platelet receptors, while both factors bind equally well to phospholipids. It seems important that factor X would not compete with X_a, because, in vivo, probably only a fraction of factor X is converted to X_a, and the binding of this X_a to the platelet must occur in the presence of a large excess of X.

Further experiments indicate that factor X_a bound to the platelet surface is quite different from X_a in solution. In the presence of heparin, antithrombin III rapidly inactivates several coagulation factors, including factor X_a (38). Once X_a is bound to the platelet, it can no longer be inactivated by the heparin/antithrombin III complex. This suggests that clotting factors activated on the platelet surface would not be inhibited by anticoagulant therapy with heparin. New interactions might be prevented, but those in progress would be expected to continue. Whether or not this explains the apparent lack of efficacy of heparin therapy in treatment of arterial thrombotic disease is an open question.

The nature of the factor X_a–binding site has been partially elucidated; the characteristics of the binding suggested a relationship to factor V for the following reasons: (*a*) Factor V also binds X_a in solution after treatment with low thrombin concentrations, and the binding requires calcium (39); (*b*) Factor V markedly increases the catalytic activity of X_a; (*c*) It has been recently shown that factor V coagulant activity of a very high specific activity appears when platelets are stimulated or broken by freezing and thawing (40); (*d*) The small number of binding sites per platelet (200 per cell) suggested that the interaction was not with a negatively charged lipid surface alone (since an equivalent lipid surface area bound millions of X_a molecules); and (*e*) Additional experiments showed that high concentrations of thrombin (50 U/ml) destroyed the X_a-binding sites in a time-dependent manner, which suggests their protein nature.

In preliminary experiments, we utilized an IgG fraction of plasma from a patient with an acquired inhibitor of factor V. This IgG fraction, when incubated with platelets after they had been treated with thrombin, blocked both subsequent X_a binding and thrombin generation; control IgG had no similar effect. When the antibody fraction was incubated with factor X_a first and then the mixture added to platelets, there was relatively little inhibition. The results indicated that the antibody was directed against something on the platelet surface, presumably factor V, and that this molecule was at least part of the factor X_a–binding site. That phospholipid may also be involved in the interaction of factor X_a, platelets, and prothrombin is suggested by an experiment of J. Stenflo (unpublished observations). He showed that the generation of thrombin by X_a bound to bovine platelets depended on the presence of γ-carboxyglutamyl residues on the prothrombin added. When abnormal prothrombin (from coumarin-treated cows) was used in place of native prothrombin, accelerated thrombin formation did not occur. Since the γ-carboxyglutamyl residues were required for binding to phospholipids, but not for accelerated thrombin formation by X_a and V alone, the experiment indicated a role for phospholipid in the platelet surface thrombin generation. The fact that rapid platelet surface thrombin formation did not occur with coumarin prothrombin was also consistent with the potent anticoagulant action of coumarin. If rapid thrombin formation had been observed, the physiological significance of platelet surface thrombin formation would have been questionable.

We suggest that the X_a receptor functions in hemostasis by localizing small amounts of factor X_a on the platelet surface at sites of platelet aggregation and release. This bound X_a then sparks intrinsic coagulation by catalyzing the rapid generation of thrombin on the platelet surface in an environment protected from plasma coagulation factor inhibitors. The regulation of this marked acceleration of thrombin generation on the platelet surface and its inhibition remains to be elucidated. Whether the factor X_a that binds to the platelet surface is actually formed on the platelets in a similar complex reaction between factors IX_a and X binding to a separate receptor site also remains to be elucidated. Similar questions can be raised about the activation of factor IX by factor XI_a. The availability of homogeneous preparations of factors XI_a and IX_a should allow these questions to be answered shortly. From past experiments in which clotting assays were used to study the interactions of factors XII_a and XI_a with platelets, Walsh (32) suggested that both of these substances could be activated and sequestered by platelets, and he postulated that by this mechanism the intrinsic coagulation system was activated by platelets. In fact, he was able to show that factor XI_a, when associated with platelets, is protected from antithrombin III inhibition in a manner somewhat similar to that found for factor X_a. These experiments suggest that activation of the intrinsic coagulation system occurs on the platelet surface by interactions that are highly specific for each of the coagulation factors and that are not mimicked by reactions occurring on model phospholipid surfaces.

Literature Cited

1. Hamberg, M., Svensson, J., Samuelsson, F. 1975. Thromboxanes: a new group of biologically active compounds derived from prostaglandin endoperoxides. *Proc. Natl. Acad. Sci. USA* 72:2994–98
2. Bills, T. K., Smith, J. B., Silver, M. J. 1976. Metabolism of [^{14}C] arachidonic acid by human platelets. *Biochim. Biophys. Acta* 424:303–14
3. Roth, G. J., Majerus, P. W. 1975. The mechanism of the effect of aspirin on human platelets. I. Acetylation of a particulate fraction protein. *J. Clin. Invest.* 56:624–32
4. Roth, G. J., Stanford, N. L., Majerus, P. W. 1975. The acetylation of prostaglandin synthetase by aspirin. *Proc. Natl. Acad. Sci. USA* 72:3073–76
5. Vane, J. R. 1971. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nature New Biol.* 231:232
6. Smith, J. B., Willis, A. L. 1971. Aspirin selectively inhibits prostaglandin production in human platelets. *Nature New Biol.* 231:235
7. Weiss, H. S. 1975. Platelet physiology and abnormalities of platelet function. *N. Engl. J. Med.* 293:531–41
8. Doolittle, R. F., Takagi, T., Cottrell, B. A. 1974. Platelet and plasma fibrinogens are identical gene products. *Science* 185:368–70
9. Ross, R., Glomset, J., Kariya, B., Harker, L. 1974. A platelet-dependent serum factor that stimulates the proliferation of arterial smooth muscle cells in vitro. *Proc. Natl. Acad. Sci. USA* 71:1207–10
10. Antoniades, H. N., Scher, C. D. 1977. Radioimmunoassay of a human serum growth factor for Balb/c-3T3 cells: Derivation from platelets. *Proc. Natl. Acad. Sci. USA* 74:1973–77
11. Majerus, P. W., Tollefsen, D. M., Shuman, M. A. 1976. The interaction of platelets with thrombin. In *Platelets in Biology and Pathology*, ed. J. L. Gordon, B. V., pp. 241–60. Amsterdam: Elsevier North Holland Biomed. Press
12. Shuman, M. A., Majerus, P. W. 1976. The measurement of thrombin in clotting blood by radioimmunoassay. *J. Clin. Invest.* 58: 1249–58
13. Tollefsen, D., Feagler, J. R., Majerus, P. W. 1974. The binding of thrombin to the surface of human platelets. *J. Biol. Chem.* 249:2646–51
14. Ganguly, P. 1974. Binding of thrombin to human platelets. *Nature* 247:306–7
15. Martin, B. M., Wasiewski, W. W., Fenton, J. W. II, Detwiler, T. C. 1976. Equilibrium binding of thrombin to platelets. *Biochemistry* 15:4886–93
16. Shuman, M. A., Majerus, P. W. 1975. The perturbation of thrombin binding to human platelets by anions. *J. Clin. Invest.* 56:945–50
17. Tollefsen, D. M., Majerus, P. W. 1976. Evidence for a single class of thrombin binding sites on human platelets. *Biochemistry* 15:2144–49
18. Detwiler, T. C., Feinman, R. D. 1973. Kinetics of the thrombin-induced release of calcium (II) by platelets. *Biochemistry* 12:282–89
19. Kahn, C. R. 1974. Membrane receptors for polypeptide hormones. *Methods Membr. Biol.* 3:81
20. Baenziger, N. L., Brodie, G. N., Majerus, P. W. 1972. Isolation and properties of a thrombin-sensitive protein of human platelets. *J. Biol. Chem.* 247:2723–31
21. Phillips, D. R., Agin, P. P. 1974. Thrombin substrates and the proteolytic site of thrombin action on human-platelet plasma membranes. *Biochim. Biophys. Acta* 352:218–27

22. Davey, M. G., Lüscher, E. F. 1967. Actions of thrombin and other coagulant and proteolytic enzymes on blood platelets. *Nature* 216:857–58
23. Tollefsen, D., Jackson, C. M., Majerus, P. P. 1975. Binding of the products of prothrombin activation to human platelets. *J. Clin. Invest.* 56:241–45
24. Chari, I. F., Feinman, R. D., Detwiler, T. C. 1977. Interrelations of platelet aggregation and secretion. *J. Clin. Invest.* 60: In press
25. Shuman, M. A., Majerus, P. . 1976. The binding of human and bovine thrombin to platelets. *Blood* 47:43–54
26. Lorand, L. 1976. Proteolytic Enzymes, Part B. *Methods Enzymol.*, 45:31–286
27. Davie, E. W., Ratnoff, O. D. 1964. Waterfall sequence for intrinsic blood clotting. *Science* 145:1310–11
28. MacFarlane, R. G. 1964. An enzyme cascade in the blood clotting mechanism and its function as a biochemical amplifier. *Nature* 202:498–99
29. Radcliffe, R., Nemerson, Y. 1976. Mechanism of activation of bovine factor VII products of change by factor X_a. *J. Biol. Chem.* 251:4797–4802
30. Colman, R. W. 1969. The effect of proteolytic enzymes on bonvine factor V. I. Kinetics of activation and inactivation by bovine thrombin. *Biochemistry* 8:1438–45
31. Rapaport, S. I., Schiffman, S., Patch, M. J., Ames, S. F. 1963. The importance of activation of anti-hemophilic globulin and proaccelerin by traces of thrombin in the generation of intrinsic prothrombinase activity. *Blood* 21:221–36
32. Walsh, P. N. 1974. Platelet coagulant activities and hemostasis: A hypothesis. *Blood* 43:597–605
33. Milstone, J. H. 1964. Thrombokinase as prime activator of prothrombin: historical perspectives and present status. *Fed. Proc.* 23:742–48
34. Hardisty, R. M., Hutton, R. A. 1966. Platelet aggregation and the availability of platelet factor 3. *Br. J. Haematol.* 12:764–76
35. Spaet, T. H., Cintron, J. 1965. Studies on platelet factor 3 availability. *Br. J. Haematol.* 11:269–75
36. Miletich, J. P., Jackson, C. M., Majerus, P. W. 1977. The interaction of coagulation factor X_a with human platelets. *Proc. Natl. Acad. Sci. USA.* 74:4033–36
37. Suttie, J. W., Jackson, C. M. 1977. Prothrombin structure, activation and biosynthesis. *Physiol. Rev.* 57:1
38. Rosenberg, R. D., Damus, P. S. 1973. The purification and mechanism of action of human antithrombin heparin cofactors. *J. Biol. Chem.* 248:6490–6505
39. Freeman, J. P., Guillin, M. C., Bezeaud, A., Jackson, C. M. 1977. Activation of bovine blood coagulation factor V, a prerequisite for it to bind both prothrombin and factor X_a. *Fed. Proc.* 36:675
40. Østerud, B., Rapaport, S. I., Lavine, K. 1977. The factor V activity of platelets: Evidence for an activated factor V molecule and for a platelet activator. *Blood* 49:819–34

Ann. Rev. Med. 1978. 29:51–66
Copyright © 1978 by Annual Reviews Inc. All rights reserved

ERYTHROPOIETIN AND THE CONTROL OF RED CELL PRODUCTION

◆7265

Stanley E. Graber, M.D. and Sanford B. Krantz, M.D.
Departments of Medicine, Veterans Administration Hospital and Vanderbilt
University School of Medicine, Nashville, Tennessee 37203

INTRODUCTION

Under normal conditions the circulating red cell mass in adults is remarkably constant. This observation suggested early on that a negative feedback mechanism was involved in the control of erythropoiesis. Furthermore, perturbation of the system by exposure to high altitude (hypoxia) resulted in increased erythropoiesis, while hypertransfusion (hyperoxia) resulted in reduced erythropoiesis. In the early 1950s, several investigators (1, 2) directly demonstrated the existence of a humoral erythropoietic factor by injecting large amounts of plasma from anemic animals into normal animals and observing an increase in blood reticulocytes or ^{59}Fe incorporation into new red cells. This humoral factor was named erythropoietin (Ep), since it did not appear to affect white cell or platelet production. During the past two decades the biochemistry and physiology of Ep have been extensively studied. It is now well established that Ep is a glycoprotein produced, at least in part, by the kidney, and that Ep is the principal factor initiating and regulating red cell production in vertebrates. Several reviews (3, 4) of the hormone have been published over the past few years. The aim of the present review is to provide the reader with a general background of Ep biology, and to update our knowledge of the hormone to the present time with particular attention directed to its mechanism of action and role in clinical medicine.

PURIFICATION AND CHEMISTRY OF ERYTHROPOIETIN

The initial work on purification of Ep was performed with plasma from anemic sheep, since relatively large volumes could be obtained and the presence of the hormone in urine was uncertain. Such plasma had a potency of 0.007 units/mg of protein, and after an extensive sequence of purification procedures, Ep with a

0066-4219/78/0401-0051$01.00

potency of 8250 units/mg was isolated (5). However, the final yield was only 0.4%. While the purification of Ep from plasma was proceeding, it became evident that urine of anemic patients provided a source of hormone with much less contaminating protein. Espada & Gutnisky (6) were able to purify human urinary Ep to 8086 units/mg of protein. Their recovery was 18.8%, or more than 50 times greater than that with plasma Ep. Miyake, Kung & Goldwasser (7) have since applied the methodology they developed for purifying sheep plasma Ep to human urinary preparations. By a combination of treatment with DEAE-cellulose, phenol-*p*-aminosalicylate, ethanol precipitation, DEAE-agarose, sulfopropyl-Sephadex, gel permeation, and hydroxylapatite chromatography, they obtained Ep at a potency of 70,400 units/mg of protein with a 21% yield.

Characterization of the purified sheep plasma Ep (8250 units/mg) showed that it was a glycoprotein that contained 24% carbohydrate. This was estimated to consist of 10% sialic acid, 4% mannose, 6% galactose, 2% glucose, and 4% N-acetyl glucosamine (8). Analysis of the highly pure human urinary Ep revealed 29% carbohydrate, which consisted of 7.5% sialic acid, 13% total hexose, and 8.9% N-acetyl glucosamine (9). The molecular weight of the human material has ranged from 27,000–60,000 as determined by gel filtration or sodium dodecyl sulfate polyacrylamide gel electrophoresis, while similar analyses of the sheep plasma material have revealed values of 46,000–61,800 daltons (8).

Despite the lack of knowledge concerning the precise chemical structure of Ep, it is clear that the sialic acid residues are necessary for biologic in vivo activity. Although asialo-Ep has full activity in vitro, it is inactive in vivo (8). Apparently the asialo form of the hormone, like most asialated glycoproteins, is rapidly cleared by the liver. This results from a specific interaction between newly exposed galactose residues on the desialated Ep and hepatic cells. Oxidation of the galactose residues, or simultaneous administration of asialo-orosomucoid, partially restores the in vivo activity of the asialated Ep (10).

The principal difficulties with the purification of Ep have been the lack of a rich source, a cumbersome bioassay, the hormone's apparent heterogeneity and instability, and the multiple time-consuming procedures necessary to produce a small amount of pure material. The initial collection and purification procedures to preserve an intact molecule and produce a high yield are critical. Recently, Sieber (11) and Spivak (12) reported the capacity to purify Ep with agarose-linked phytohemagglutinin or wheat germ agglutinin. A 7- to 50-fold purification was obtained with a 73–95% yield of the hormone using this relatively simple procedure. Utilization of lectin chromatography should enhance the yield and ease purification of the hormone. However, pure Ep will continue to be in short supply because of its limited source.

ASSAY

The current assay (13) utilizes either hypertransfused or exhypoxic mice. The above-normal red cell concentration produces an almost complete cessation of erythropoiesis. Ep or a test substance is then administered and this is followed in one to three days by parenteral ^{59}Fe. Three days later the animals are bled and the amount of ^{59}Fe in the red cells determined. This system can detect small amounts of the hormone (0.05 units) and has a 40-fold range of response, but it cannot measure the amount of hormone in normal serum. For normal individuals and patients with reduced or only slightly increased amounts of Ep, concentrated urine specimens must be assayed to measure the level of hormone. This assay also suffers from a lack of precision, is time consuming, and requires the use of large numbers of mice per group. In addition, androgens, cobalt, certain vasoactive amines (such as histamine, tryptomine, and serotonin), prostaglandins, cyclic nucleotides, and agents that cause hemolysis may increase erythropoiesis and produce false-positive results. Nevertheless, with proper controls this is overall the most reliable and widely used assay available.

The standard unit of measure for Ep is based on the erythropoietic effect of cobalt. Injection of cobalt into starved rats produces a dose-related increase in erythropoiesis identical to that produced by crude Ep. In order to provide for a consistent, easily obtainable standard, one unit of Ep was set equal to the response of 5 mmoles of cobalt in the fasted-rat assay. When sheep Ep extracts became available, one batch was set aside as Standard A. One unit of Standard A was 0.05 mg and it produced a response equivalent to one cobalt unit. As this material was used up it was replaced by Standard B, a lyophilized crude preparation of human urine Ep. The international unit for Ep is currently 1.48 mg of Standard B, which produces a response similar to that observed with one cobalt unit or one unit of Standard A (14).

Tissue culture assays have been developed that are very sensitive to Ep and give excellent dose response curves. These systems utilize either rat marrow (15) or fetal liver cells (16) and are based on incorporation of ^{59}Fe into heme. Amounts of Ep as small as 0.001 to 0.005 units can be measured. Recently, a plasma clot culture technique (17) using mouse or human marrow cells (18) has been described in which erythroid colonies form in vitro. This system also responds to Ep in a dose-related fashion and could be used to assay the hormone. These methods, however, may detect asialo-Ep (15), which is inactive in vivo, and may be sensitive to other factors in serum and urine, making it difficult to determine the amount of Ep in crude preparations. They have been of value in following the Ep concentration in diverse fractions produced by purification procedures and in doing stability studies (5, 7), but ultimately the polycythemic mouse assay must be used for definitive measurement of the hormone.

Several immunologic assays have been proposed for Ep including a hemagglutination inhibition assay and a radioimmunoassay. The hemagglutination inhibition system has been commercially marketed, but has many serious problems, which appear to be related to the specificity of the antiserum (19). Two different laboratories (20, 21) have developed radioimmunoassays, which utilize highly purified Ep preparations (8000–12,000 units/mg) labeled with ^{125}I. The assays appear to be sufficiently sensitive to quantitate normal serum Ep levels, but the range of normal found in the two laboratories varied considerably. Both groups observed that patients with renal failure had very high serum Ep levels by radioimmunoassay even though several of the same patients had no detectable activity by the polycythemic mouse assay. In addition, it was shown that asialo-Ep was immunoreactive, but inactive in vivo. These studies suggested that serum from patients with renal failure might contain an immunoreactive but biologically inactive Ep. This is a very interesting hypothesis and, if confirmed, may increase our understanding of the anemia of renal disease. Much more correlation of the bioassay and radioimmunoassay in diverse clinical situations is needed before the latter can be used for routine assay of the hormone.

PRODUCTION OF ERYTHROPOIETIN

The Kidney

In 1957, Jacobson et al (22) performed bilateral nephrectomies on rats and showed that these animals no longer responded to phlebotomy or cobalt administration with increased Ep production. Rats that only had their ureters ligated and that developed similar elevations of the blood urea nitrogen still responded to cobalt or phlebotomy in the usual fashion. Similar results were found in mice, rabbits, baboons, and dogs. In humans, Ep levels were found to be low in the serum of anemic patients with severe uremia, but after successful renal transplantation increases in blood Ep levels and hematocrit were observed (23). Furthermore, erythropoietic activity was detected in the perfusates from isolated rabbit kidneys (24) and was produced by renal cell cultures in vitro (25). Although the evidence to support the renal origin of Ep was substantial, Ep had not been extracted from normal kidneys or from various normal kidney fractions until very recently (26). To explain this difficulty, three different hypotheses have been proposed: (*a*) the erythrogenin theory, (*b*) the pro-Ep theory, and (*c*) the inhibitor theory.

The erythrogenin theory states that the kidney, in response to hypoxia, produces a substance that enzymatically converts a circulating α-globulin to active Ep. The renal erythropoietic factor is called erythrogenin and the plasma substrate, which is thought to be of hepatic origin, is called erythropoietinogen. Such a factor has been extracted from the light mitochondrial fraction of kidneys and is capable of generating Ep when incubated with plasma from normal animals. This factor has little direct effect on erythropoiesis in plethoric mice, and the kinetics of its reaction with serum suggests it may be an enzyme (27). This work has been reproduced in some laboratories (28), but not in others (29).

According to the proerythropoietin theory, the kidneys produce an inactive molecule which contains intact Ep. Pro-Ep is then activated by a serum factor to produce biologically active Ep. Evidence for this consists of the following: (*a*) Incubation of a light mitochondrial fraction from hypoxic rat kidneys with serum from normal rabbits results in active Ep which is immunologically rat in origin (30). (*b*) Antibody to Ep labeled with fluorescein binds to renal glomeruli, which indicates the presence of Ep in the kidneys (31).

The inhibitor theory suggests that complete Ep is produced by the kidney, but it may be inactivated by a lipid-soluble inhibitor during the extraction process.

Contact with normal serum neutralizes the inhibitor and allows extraction of biologically active Ep (32). Further support for this, and also the pro-Ep theory, comes from the observation that when hourly Ep production in an intact hypoxic animal was compared to a similarly stimulated kidney perfused in vitro with a medium almost completely free of serum, production of Ep continued, which indicates that the kidneys were capable of producing the hormone without need for much serum protein (33).

A consensus does not yet exist on how Ep is made by the kidney. Much more work is necessary to determine which of the above hypotheses might be correct. However, it seems clear that the kidney is the principal site regulating production of the hormone.

Role of Prostaglandins and Cyclic AMP in Production of Ep

Current evidence indicates that the stimulation of Ep production by hypoxia or ischemia may be mediated via increases in renal prostaglandin and cyclic AMP levels. The administration of cyclic AMP to plethoric mice increased Ep production (34) and led to an increased red cell mass in normal mice (35). Prostaglandin E_1 raised the renal cyclic AMP concentration in isolated perfused dog kidneys as well as the amount of Ep in the perfusates (34); the administration of prostaglandin E_2 to polycythemic mice increased Ep levels and red blood cell production (36). Renal artery constriction increased both prostaglandin E and Ep blood levels in dogs. The effect on both these parameters as well as on hypoxia-induced Ep production was blocked by indomethacin, a known inhibitor of prostaglandin synthesis (37). Since prostaglandins are located mainly in the renal medulla, while Ep has been identified by fluorescent antibodies in the renal cortex (31), these observations are most consistent with the hypothesis that renal hypoxia or ischemia produces a medullary release of prostaglandin E_2 which in turn elevates renal cortical cyclic AMP levels. The newly formed cyclic AMP then may activate a protein kinase that leads to increased Ep production, possibly through phosphorylation of a precursor renal erythropoietic factor (38).

Extrarenal Erythropoietin

Although the kidney is the major site controlling Ep formation, after bilateral nephrectomy the production of hormone and erythropoiesis continues in most species, but at a considerably lower level. Anephric mice and rats retain about 10% of their original capacity to increase Ep production in the presence of hypoxia (39), while anephric patients who become very anemic have Ep in their plasma (40). The extrarenal Ep of rats appears to be similar to Ep of renal origin since it stimulates heme synthesis by marrow cells and is neutralized by antiserum to urinary Ep (41). Several lines of study point to the liver as the primary site of extrarenal Ep production. Ep is produced primarily in the liver during fetal life (42). Hepatectomy abolishes hypoxia-stimulated hormone production in anephric rats (43), but not in animals with intact kidneys. In partially hepatectomized rats, the extrarenal Ep response to hypoxia correlates directly with liver regeneration and is most increased during the period of greatest proliferation (44). Administration of colloidal carbon or zymosan, which induces liver reticuloendothelial cell hyperplasia, enhances Ep production in nephrectomized rats exposed to hypoxia. This effect, which is also blocked by hepatectomy, suggests that the liver reticuloendothelial system (Kupffer cells) may be an important factor in extrarenal Ep production (45).

SITE AND MECHANISM OF ACTION

Effect on Precursors of Erythroblasts

Ep acts on the bone marrow to selectively increase erythropoiesis, while not affecting white cell or platelet production. Jacobson et al (46) and later Filmanowicz & Gurney (47) provided evidence that the hormone is directly involved in the initiation of normal red cell differentiation from a primitive, morphologically undifferentiated cell. They injected hypertransfused mice, whose marrow and spleen contained virtually no recognizable red cell precursors, with a single dose of Ep. Proerythroblasts were evident at 24 hr and by 72 hr maturation had proceeded to erythroblasts and reticulocytes. When Ep stimulation was quickly terminated with an anti-Ep antibody, the wave of erythroid maturation continued, which indicated that Ep was not necessary for the continued maturation of the erythroblasts. In the early 1960s Till & McCulloch (48) developed a technique for assaying a hematopoietic spleen colony–forming unit (CFU-S). This unit was subsequently shown to be a pluripotential hematopoietic stem cell from which erythrocytic, granulocytic, and megakaryocytic cell lines originated. Considerable evidence suggested that Ep did not act on the CFU-S, but instead acted on more developed descendents of the CFU-S designated the erythropoietin-responsive cells (ERC): (*a*) Stimulated differentiation of ERC by Ep did not produce a decline in CFU-S (49). (*b*) A response to Ep still occurred in animals with a lack of CFU-S due to irradiation or drug treatment (50). (*c*) ERC were destroyed by large doses of tritiated thymidine, which had little effect on CFU-S (51). (*d*) CFU-S were separated from ERC by unit gravity sedimentation (52). (*e*) The number of erythroid and granulocytic colonies grown simultaneously in solid culture were independent of each other (53).

Characterization of the ERC has been difficult as the cells making up this compartment are morphologically undefined. Indirect experiments utilizing irradiation and busulfan to prevent the formation of new ERC from CFU-S, and actinomycin D to suppress the formation of erythroblasts, suggested that Ep acts on the ERC to stimulate proliferation of these cells (54). Further characterization was facilitated by the development of plasma clot (53) and methylcellulose (55, 56) culture techniques capable of supporting growth of discrete erythroid colonies. With these methods, at least two different ERC have been described in mouse marrow (56, 57). These cells appear to be distinct from recognizable erythroblasts and have been named the erythrocytic colony–forming unit (CFU-E) and erythrocytic burst–forming unit (BFU-E). They can be separated by unit gravity sedimentation (58), and the number of colonies formed from each is dependent upon and proportional to the concentration of Ep present and the number of marrow cells plated (56, 59). CFU-E give rise to colonies that consist of a single group of 8–32 erythroblasts, occur in maximum number 48 hr after initiation of the cultures, and require small amounts of Ep (0.01 to 0.25 units) for formation. BFU-E give rise to grouped clusters of erythroblasts that may number several hundred to several thousand cells (bursts). They occur in maximum number after day 7 or 8 of culture, and require large amounts of Ep (0.1 to 2.5 units). Moreover, CFU-E are increased by bleeding, depressed by hypertransfusion, and very sensitive (75% destroyed) to thymidine suicide (57, 60). In contrast, BFU-E are little changed by bleeding or hypertransfusion, resistant (30% destroyed) to thymidine suicide, and quantitatively correlate well with the number of CFU-S in a given specimen (57, 59, 60). In addition, other precursor cells with characteristics intermediate between the BFU-E and CFU-E have been described (59) and BFU-E have been found in the blood and marrow of humans, while CFU-E have been found in the marrow, but not the blood (18, 61). These data are compatible with the notion that the BFU-E is a very early cell close to the pluripotential stem cell (CFU-S) whose generation is not controlled by Ep, but which responds to Ep with reduced sensitivity to give rise (possibly through several intermediates) to the CFU-E. The CFU-E is a later cell, close to the earliest recognizable erythroblast, whose generation and differentiation into erythroblasts are controlled by Ep.

Effect on Erythroblasts

In addition to acting on the ERC to initiate development into proerythroblasts, Ep has also been shown to act on early erythroblasts: (*a*) Ep increased proliferation of cell populations highly enriched for proerythroblasts and basophilic erythroblasts (62). (*b*) Through the use of autoradiography, Ep was shown to increase RNA synthesis in early fetal liver erythroid cells (63). (*c*) An Ep concentrate increased the blood reticulocyte count within 24 hr, which was too short a time for development from the CFU-E (64). (*d*) Ep produced a decreased marrow transit time of erythroid cells with early denucleation of erythroblasts and possible bypass of the orthochromatic stage (65). (*e*) When incubated with normal mouse marrow cells, Ep increased heme synthesis within 4 hr, but when incubated with marrow from hypertransfused mice, which had virtually no erythroblasts, the increase in heme synthesis was delayed until 15–24 hr after addition of the hormone (66).

Effect on RNA and DNA Synthesis

The mechanism by which Ep increases red cell production has received considerable attention. Krantz & Goldwasser (67) showed that bone marrow RNA synthesis was stimulated within 15 min after the addition of Ep. Several investigators have confirmed this RNA effect and have further characterized the different types of RNA stimulated (63, 66, 68–70). It was found that the early increase consisted mostly of a very large species (150S) that was present in only small amounts in unstimulated cells. At later times (1–4 hr), Ep stimulated the synthesis of rapidly labeled RNA fractions sedimenting at 55 to 65, 45, 9, 6, and 4S. Specifically, increases in ribosomal precursor RNA (45S), ribosomal RNA (28 and 18S), and transfer RNA were demonstrated. The stimulation of the 9S fraction was thought to represent formation of globin messenger, as reticulocyte messenger RNA had been shown to sediment at 9S (71). Ep-induced formation of globin messenger RNA was then directly demonstrated in mouse fetal liver cells by both complementary DNA hybridization (72) and initiation of globin synthesis in a cell-free system (73). Small increases in globin messenger were found within 6 hr after addition of the hormone, while a 250-fold rise was present by 22 hr. Recent studies with exhypoxic mouse spleen have shown that the activity of RNA polymerase II, the enzyme thought to synthesize messenger RNA, is increased 30 min after administration of Ep. This is followed by increases in RNA polymerase I activity and nonhistone protein synthesis at 3 hr, and by increased histone synthesis at 18 hr (74, 75).

Increases in DNA synthesis were also shown to occur, but generally followed the increase in RNA synthesis (76, 77). Moreover, inhibition of DNA synthesis did not block Ep-induced stimulation of RNA formation (63, 77), while treatment with actinomycin D (which blocked RNA synthesis) obliterated increased DNA formation, ^{59}Fe uptake, and hemoglobin synthesis (68, 77, 78). Thus, it appears that stimulation of RNA synthesis is the primary nuclear event in the molecular action of Ep. This is followed by DNA synthesis, cell division, and maturation (hemoglobin production) in the responsive cells.

Effect on a Second Messenger

Ep may stimulate RNA synthesis either directly by entering the cell and going to the nucleus, or indirectly through an intermediate intracellular substance, i.e. a second messenger. The rapidity of the RNA effect, the large molecular weight of the hormone, the lack of consumption of the hormone in tissue culture (79), and the reversible obliteration of Ep responsiveness in trypsin-treated cells (80) all suggest that Ep may be interacting with the cell membrane to produce such a second messenger. Cyclic AMP has been an obvious candidate for this role. However, the evidence has been contradictory. Cyclic AMP was shown to increase heme synthesis and erythroid colony formation in cultures of marrow cells from several species (81), but Ep did not affect cyclic AMP levels in rat fetal liver cells, a tissue very sensitive to the hormone (82). The work of Brown & Adamson (83) suggested an explanation for this apparent paradox. Utilizing the plasma clot technique, they found that, although at a fixed dose of hormone cyclic AMP increased the number of erythroid colonies, only Ep was capable of initiating the growth of such colonies. Moreover, cyclic AMP responsive cells could be separated from Ep responsive cells by unit gravity sedimentation. Thus, it appears that the action of Ep is not mediated by cyclic AMP, but that cyclic AMP may act on an erythroid precursor cell to increase its sensitivity to Ep. Further studies (84) have indicated that these cells have receptors for β-adrenergic agents and the latter may modulate erythropoiesis via this cyclic AMP mechanism.

Cyclic GMP is another attractive candidate for a second messenger. This nucleotide rises in lymphocytes stimulated with phytohemagglutinin or concanavalin A and in resting mouse fibroblasts stimulated by serum (85, 86). These findings have suggested that an increase in cyclic GMP levels may serve as a mitogenic signal. Moreover, Ep produces a marked increase in the cyclic GMP concentration in rabbit marrow and fetal liver cells (87, 88). However, this increase does not occur until at least 1.5 hr after addition of the hormone (88). Since Ep is known to increase RNA synthesis in fetal liver and marrow cells within minutes, cyclic GMP does not appear to represent a messenger in that process, but may be involved in a later aspect of the hormone's effect.

A third candidate for a second messenger has recently been described by Chang & Goldwasser (89). They extracted a cytoplasmic fraction from Ep-treated rat marrow cells that stimulated RNA synthesis in marrow cell nuclei. A similar cytoplasmic fraction from Ep-treated kidney and lung cells, and Ep itself, were inactive. The production of an active fraction appeared to depend on the presence of marrow cell cytoplasmic membrane or intact cells, since the incubation of Ep with normal marrow cell cytosol did not lead to active material. The factor appeared promptly after addition of Ep and peaked within 1 hr. Although inactivation by trypsin suggested that it had a peptide structure, inhibition of protein synthesis did not block its appearance. These experiments suggest that Ep has its primary effect on the cytoplasmic membrane of marrow cells to produce an active cytoplasmic protein intermediate that interacts with the nucleus to stimulate synthesis of a variety of RNAs.

MEDICAL IMPLICATIONS

Anemia

Anemia is due to loss of red cells by hemorrhage or hemolysis, or to decreased red cell production by the marrow. Excessive destruction or loss of red cells is accompanied by increased levels of Ep and increased erythropoiesis. Marrow production defects can be further subdivided into those with either increased or inappropriately low Ep levels. In the former, represented by aplastic anemia and B_{12}, folate, or iron deficiency, the anemia is due to a primary injury or deficiency of the marrow, and the increased Ep that results from the anemia is unable to establish normal erythropoiesis. In the latter, the lack of an increase in Ep to a degree consistent with the degree of anemia indicates that a reduced supply of Ep may be partially responsible for the anemia. Specifically, two thirds of anemic patients with rheumatoid arthritis, chronic infection, and some malignancies did not have increased Ep levels despite the anemia (90). One third of these patients had mild elevations of their serum hormone levels, but the increase was extremely small for the degree of anemia. Injection of bacterial neuraminidase into mice has been shown to inactivate Ep in vivo (91). Increases of this enzyme might play a role in producing the anemia in some of the above conditions, but further experiments are needed to test this idea. The anemia of protein deprivation is probably due at least in part to reduced Ep production, since protein deprivation leads to an anemia in laboratory animals characterized by low Ep levels and complete correction with small amounts of exogenous hormone (92). Endocrine deficiencies produced in animals by castration, hypophysectomy, or thyroidectomy also elicit an anemia due to reduced Ep production, as well as a reduced stimulation of the marrow cells by the primary hormone (93).

In humans the anemia attracting the most attention in terms of its possible relation to a lack of Ep has been the anemia of renal disease. Chronic renal disease with progressive loss of renal mass appears to result in decreased production of biologically active Ep and an anemia that is generally proportional to the degree of renal damage. By the polycythemic mouse assay, most uremic patients have inappropriately low blood or urinary Ep levels, but following renal transplantation elevated levels of Ep are found and normal erythropoiesis ensues (23). Anephric humans generally have a more severe anemia than comparable dialysis patients who have not undergone nephrectomy (94), and both uremic animals and a few humans have been shown to respond to serum rich in Ep or crude Ep preparations with increased erythropoiesis (95, 96). Besides the relative lack of Ep, other factors such as hemorrhage, hemolysis, increased red cell 2,3-diphosphoglycerate levels (97), and the accumulation of toxic metabolites (98) may play a role in the pathogenesis of the anemia of renal disease. From the limited studies available it appears that more Ep is necessary to stimulate erythropoiesis in a uremic human than a normal individual (99). Currently, androgens are used to treat the anemia of chronic renal failure; they produce an increase in Ep levels and the red cell mass in most hemodialyzed patients who have not undergone nephrectomy (94). Anephric patients, on the other hand, usually respond either poorly or not at all. Because androgens produce many undesirable side effects, Ep may be a better therapeutic agent for the treatment of the anemia of uremia. Chronic administration of purified Ep might be effective in raising the red cell mass. However, no source of purified hormone is presently available and current impure preparations are clearly not suited for chronic administration to patients.

Polycythemia

The discovery of Ep and its regulation of red cell production has led to a more meaningful classification of polycythemia. In most cases the increased erythropoiesis is an appropriate physiologic response to hypoxia which increases Ep production. The hypoxia can arise from the atmosphere, pulmonary or vascular disease, a reduced oxygen capacity of the blood, or a reduced capacity to release oxygen from the blood. The latter occurs most frequently with an abnormal hemoglobin, an inherited autosomal dominant condition (100), but also with reduced red cell 2,3-diphosphoglycerate, an inherited autosomal recessive condition (101). An increase in urinary Ep is generally found. In some patients with hypoxic polycythemia, the urinary level is not increased, but in these patients phlebotomy to a normal hematocrit results in an exaggerated increase in the daily urinary excretion of the hormone (102). It is possible in these cases that the bioassay for Ep is not sensitive enough to measure small increases in basal Ep excretion and that the patients have partially compensated for the hypoxia in another manner.

In other cases of polycythemia an inappropriate increase in Ep production is present, since no physiologic hypoxic stimulus can be identified. The increased Ep in these patients appears to arise autonomously and is associated predominantly with renal lesions such as cysts, hydronephrosis, adenomas, and carcinomas, but also with cerebellar hemangioblastomas, hepatomas, uterine leiomyomas, and a variety of tumors and hyperplasias of the endocrine organs. Ep has been extracted from some of these tumors, and phlebotomy results in no further increase in the urinary Ep levels (102, 103). A selective increase in the red cell concentration is a primary clinical characteristic of polycythemia secondary to increased Ep production, since the white cell and platelet counts are unaffected by the hormone. Two families with an inherited recessive polycythemia have been found to have an inappropriate increase in Ep production, which is due to a cellular defect at the site of Ep synthesis (104). No increase in urinary Ep was seen in either family in response to phlebotomy.

In polycythemia vera an inappropriate increase in erythropoiesis occurs without increased Ep production. Urinary Ep levels are depressed and phlebotomy induces only a modest increase in urinary Ep (102, 103). The white cell and platelet counts are often increased, since the disease is not limited to abnormal red cell production. Short-term cultures of marrow cells from patients with polycythemia vera indicated that the Ep-stimulated increase in hemoglobin or DNA synthesis was quite small when compared with normal marrow cells, but returned to normal once the patients were put into remission with busulfan therapy (105, 106). These experiments were consistent with the hypothesis that two cell lines were present; an abnormal clone that was relatively insensitive to Ep and that proliferated autonomously, and a normal cell line that was regulated by Ep and expressed itself only after treatment of the disease with busulfan. This concept of the clonal nature of polycythemia vera has been further supported by recent studies in which polycythemia vera marrow from two glucose-6-phosphate dehydrogenase (G-6-PD) A,B heterozygotes was cultured in methyl cellulose (107). With this technique, the G-6-PD type could be determined for individual erythroid colonies. In the absence of added Ep, marrow erythroid colonies (endogenous colonies) appeared, whereas in normal marrow cell cultures no erythroid colonies are present without added Ep. Furthermore, all these colonies were type A. With the addition of increasing doses of Ep, an increasing number of type B as well as type A colonies appeared in vitro. These studies indicate the clonal nature of this disease, the response of the abnormal clone (type A) to Ep, and the presence of presumably normal, but suppressed, type B cells. Other experiments with the plasma clot technique and polycythemia vera marrow have confirmed the appearance of endogenous colonies and, in addition, have shown enhanced erythroid colony formation in response to Ep (108, 109). Moreover, when the culture media was treated with Ep antibody to remove virtually all the hormone, the endogenous colonies were greatly reduced in number. These latter experiments (108, 109) indicated that in polycythemia vera the size of the ERC compartment may be increased and that a fraction of these cells have an increased sensitivity to Ep. Accordingly, only very small amounts of Ep would be needed to maintain an elevated red cell mass. The discrepancy in Ep sensitivity of polycythemia vera marrow cells when cultured by different methods is still not resolved. Normal marrow cells increase their Ep responsiveness as they develop from BFU-E to CFU-E. It is possible that different stages in progenitor cell development are being assayed in the different culture methods and that polycythemia vera progenitor erythroid cells have an altered development of Ep sensitivity as compared to normal marrow cells. Nevertheless, it seems clear that polycythemia vera is a clonal disease in which a cell line develops that gives rise to new erythroid cells in the presence of only very small amounts of Ep, and in which a normal cell line under physiologic hormonal regulation is suppressed.

Literature Cited

1. Borsook, H. A., Graybiel, A., Keighley, G., Windsor, E. 1954. Polycythemic response in normal adult rats to a non-protein plasma extract from anemic rabbits. *Blood* 9:734–42
2. Plzak, L. F., Fried, W., Jacobson, L. O., Bethard, W. F. 1955. Demonstration of stimulation of erythropoiesis by plasma from anemic rats using ^{59}Fe. *J. Lab. Clin. Med.* 46:671–78
3. Krantz, S. B., Jacobson, L. O. 1970. *Erythropoietin and the Regulation of Erythropoiesis*. Chicago: Univ. Chicago Press. 330 pp.
4. Fried, W. 1975. Erythropoietin and the kidney. *Nephron* 15:327–49
5. Goldwasser, E., Kung, C. K. H. 1971. Purification of erythropoietin. *Proc. Natl. Acad. Sci. USA* 68:697–98
6. Espada, J., Gutnisky, A. 1970. Purificacion de eritropoyetina urinaria humana. *Acta Physiol. Lat. Am.* 20:122–29
7. Miyake, T., Kung, C. K. H., Goldwasser, E. 1977. The purification of human erythropoietin. *J. Biol. Chem.* 252:5558–64
8. Goldwasser, E. 1975. Erythropoietin and the differentiation of red blood cells. *Fed. Proc.* 34:2285–92
9. Espada, J., Langton, A. A., Dorado, M. 1972. Human erythropoietin. Studies on purity and partial characterization. *Biochim. Biophys. Acta* 285:427–35
10. Goldwasser, E., Kung, C. K. H., Elliason, J. 1974. On the mechanism of erythropoietin-induced differentiation. XIII. The role of sialic acid in erythropoietin action. *J. Biol. Chem.* 249:4202–6
11. Sieber, F. 1977. Chromatography of human urinary erythropoietin and granulocyte colony-stimulating factor on insolubilized phytohaemagglutinin. *Biochim. Biophys. Acta* 496:146–54
12. Spivak, J. L., Small, D., Hollenberg, M. D. 1976. *Blood* 48(6):990 (Abstr.)
13. Kazal, L. A., Erslev, A. J. 1975. The measurement of erythropoietin. *Ann. Clin. Lab. Sci.* 5:91–97
14. Krantz, S. B., Jacobson, L. O. 1970. See Ref. 3, pp. 11–12
15. Goldwasser, E., Eliason, J. F., Sikkema, D. 1975. An assay for erythropoietin *in vitro* at the milliunit level. *Endocrinology* 97:315–23
16. Dunn, C. D. R., Jarvis, J. H., Greenman, J. M. 1975. A quantitative bioassay for erythropoietin using mouse fetal liver cells. *Exp. Hematol. (Copenhagen)* 3:65–78
17. McLeod, D. L., Shreeve, M. M., Axelrad, A. A. 1974. Improved plasma culture system for production of erythrocytic colonies *in vitro:* Quantitative assay method for CFU-E. *Blood* 44:517–34
18. Tepperman, A. D., Curtis, J. E., McCulloch, E. A. 1974. Erythropoietic colonies in cultures of human marrow. *Blood* 44:659–69
19. Kolk-Vegter, A. J., Kolk, A. H. J., Napier, J. A. F., Dunn, C. D. R. 1975. Some problems concerning

the assay of erythropoietin using the haemagglutination inhibition kit. *Br. J. Haematol.* 30:371–72

20. Lertora, J. J. L., Dargon, P. A., Rege, A. B., Fisher, J. W. 1975. Studies on a radioimmunoassay for human erythropoietin. *J. Lab. Clin. Med.* 86:140–51
21. Garcia, J. F. 1974. Radioimmunoassay of human erythropoietin. In *Radioimmunoassay and Related Procedures in Medicine,* 1:275–87. Vienna: Int. Atomic Energy Agency
22. Jacobson, L. O., Goldwasser, E., Fried, W., Plzak, L. 1957 Studies on erythropoiesis. VII. The role of the kidney in the production of erythropoietin. *Trans. Assoc. Am. Physicians.* 70:305–17
23. Krantz, S. B., Jacobson, L. O. 1970. See Ref. 3, pp. 29–31
24. Kuratowska, Z., Lewartowski, B., Michulak, E. 1961. Studies on the production of erythropoietin by isolated perfused organs. *Blood* 18:527–34
25. Burlington, H., Cronkite, E. P., Reincke, U., Zanjani, E. D. 1972. Erythropoietin production in cultures of goat renal glomeruli. *Proc. Natl. Acad. Sci. USA* 69:3547–50
26. Sherwoon, J. B., Goldwasser, E. 1977. *Extraction of erythropoietin activity from kidney.* Presented at Ann. Meet. Endocrine Soc., 59th, Chicago
27. Zanjani, E. D., McLaurin, W. D., Gordon, A. S., Rappaport, I. A., Gibbs, J. M., Gidari, A. S. 1971. Biogenesis of erythropoietin: Role of the substrate for erythrogenin. *J. Lab. Clin. Med.* 77:751–58
28. Malgor, L. A., Fisher, J. W. 1970. Effects of testosterone on erythropoietin production in isolated perfused kidneys. *Am. J. Physiol.* 218:1732–36
29. Erslev, A. J., Kazal, L. A. 1969. Renal erythropoietic factor: Lack of effect on hypertransfused mice. *Blood* 34:222–29
30. Peschle, C., Condorelli, M. 1975. Biogenesis of erythropoietin: Evidence for pro-erythropoietin in a subcellular fraction of kidney. *Science* 190:910–12
31. Busuttil, R. W., Roh, B. L., Fisher, J. W. 1971. The cytological localization of erythropoietin in the human kidney using the fluorescent antibody technique. *Proc. Soc. Exp. Biol. Med.* 137:327–30
32. Erslev, A. J., Kazal, L. A., Miller, O. P. 1971. The action and neutralization of a renal lipid inhibitor of erythropoietin. *Proc. Soc. Exp. Biol. Med.* 138:1025–29
33. Erslev, A. J. 1974. *In vitro* production of erythropoietin by kidneys perfused with a serum-free solution. *Blood* 44:77–85
34. Rodgers, G. M., Fisher, J. W., George, W. J. 1975. The role of renal adenosine 3′,5′-monophosphate in the control of erythropoietin production. *Am. J. Med.* 58:31–38
35. Rodgers, G. M., Fisher, J. W., George, W. J. 1975. Increase in hematocrit, hemoglobin and red cell mass in normal mice after treatment with cyclic AMP. *Proc. Soc. Exp. Biol. Med.* 148:380–82
36. Gross, D. M., Brookins, J., Fink, G. D., Fisher, J. W. 1976. Effects of prostaglandins A_2, E_2 and $F_{2\alpha}$ on erythropoietin production. *J. Pharmacol. Exp. Ther.* 198:489–96
37. Mujovic, V. M., Fisher, J. W. 1975. The role of prostaglandins in the production of erythropoietin (ESF) by the kidney. II. Effects of indomethacin on erythropoietin production following hypoxia in dogs. *Life Sci.* 16:463–73
38. Martelo, O. J., Toro, E. F., Hirsch, J. 1976. Activation of renal erythropoietic factor by phosphorylation. *J. Lab. Clin. Med.* 87:83–88
39. Jacobson, L. O., Marks, E. K., Gaston, E. O., Goldwasser, E. 1959. Studies on erythropoiesis. XI. Reticulocyte response of transfusion-induced polycythemic mice to anemic plasma from nephrectomized mice and to plasma from nephrectomized rats exposed to low oxygen. *Blood* 14:635–43
40. Krantz, S. B., Jacobson, L. O. 1970. See Ref. 3, p. 30
41. Fried, W., Kilbridge, T., Krantz, S., McDonald, T., Lange, R. 1969. Studies on extrarenal erythropoietin. *J. Lab. Clin. Med.* 73:244–48
42. Zucali, J. R., Stevens, V., Mirand, E. A. 1975. *Invitro* production of erythropoietin by mouse fetal liver. *Blood* 46:85–90
43. Fried, W. 1972. The liver as a source of extrarenal erythroipoietin. *Blood* 40:671–77

44. Naughton, B. A., Kaplan, S. M., Roy, M., Burdowski, A. J., Gordon, A. S., Piliero, S. J. 1977. Hepatic regeneration and erythropoietin production in the rat. *Science* 196:301–2
45. Peschle, C., Marone, G., Genovese, A., Rappaport, I. A., Condorelli, M. 1976. Increased erythropoietin production in anephric rats with hyperplasia of the reticuloendothelial system induced by colloidal carbon or zymosan. *Blood* 47:325–37
46. Jacobson, L. O., Goldwasser, E., Plzak, L., Fried, W. 1957. Studies on erythrocytosis. IV. Reticulocyte response of hypophysectomized and polycythemic rodents to erythropoietin. *Proc. Soc. Exp. Biol. Med.* 94:243–49
47. Filmanowicz, E., Gurney, C. W. 1961. Studies on erythropoiesis. XVI. Response to a single dose of erythropoietin in polycythemic mouse. *J. Lab. Clin. Med.* 57:65–72
48. Till, J. E., McCulloch, E. A. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* 14:213–22
49. Bruce, W. R., McCulloch, E. A. 1964. The effect of erythropoietic stimulation on the hematopoietic colony-forming cells of mice. *Blood* 23:216–32
50. O'Grady, L. F., Lewis, J. P., Trobaugh, F. E. Jr. 1968. The effect of erythropoietin on differentiated erythroid precursors. *J. Lab. Clin. Med.* 71:693–703
51. Porteous, D. D., Lajtha, L. G. 1968. Restoration of stem cell function after irradiation. *Ann. NY Acad. Sci.* 149:151–55
52. Stephenson, J. R., Axelrad, A. A. 1971. Separation of erythropoietin-sensitive cells from hemopoietic spleen colony-forming stem cells of mouse fetal liver by unit gravity sedimentation. *Blood* 37:417–27
53. Stephenson, J. R., Axelrad, A. A., McLeod, D. L., Shreeve, M. M. 1971. Induction of colonies of hemoglobin synthesizing cells by erythropoietin *in vitro*. *Proc. Natl. Acad. Sci. USA* 68:1542–46
54. Reissman, K. R., Udupa, K. B. 1972. Effect of erythropoietin on proliferation of erythropoietin-responsive cells. *Cell Tissue Kinet.* 5:481–89
55. Iscove, N. N., Sieber, F., Winterhalter, K. H. 1974. Erythroid colony formation in cultures of mouse and human bone marrow: Analysis of the requirement for erythropoietin by gel filtration and affinity chromatography on agarose-concanavalin A. *J. Cell. Physiol.* 83:309–20
56. Iscove, N. N., Sieber, F. 1975. Erythroid progenitors in mouse bone marrow detected by macroscopic colony formation in culture. *Exp. Hematol. (Copenhagen)* 3:32–43
57. Axelrad, A. A., McLeod, D. L., Shreeve, M. M., Heath, D. S. 1973. In *Proc. Int. Workshop Hemopoiesis in Culture, 2nd,* ed. W. A. Robinson, pp. 226–37. New York: Grune & Stratton
58. Heath, D. S., Axelrad, A. A., McLeod, D. L., Shreeve, M. M. 1976. Separation of the erythropoietin-responsive progenitors BFU-E and CFU-E in mouse bone marrow by unit gravity sedimentation. *Blood* 47:777–92
59. Gregory, C. 1976. Erythropoietin sensitivity as a differentiation marker in the hemopoietic system: Studies of three erythropoietic colony responses in culture. *J. Cell. Physiol.* 89:289–302
60. Iscove, N. N. 1976. The role of erythropoietin in regulation of population size and cell cycling of early and late erythroid precursers in mouse bone marrow. *Cell Tissue Kinet.* 10:323–34
61. Clarke, B. J., Nathan, D., Hillman, D., Alter, B., Housman, D. 1976. *Blood* 48(6):980 (Abstr.)
62. Glass, J., Lavidor, L. M., Robinson, S. H. 1975. Use of cell separation and short term culture techniques to study erythroid cell development. *Blood* 46:705–11
63. Djaldetti, M., Preisler, H., Marks, P. A., Rifkind, R. A. 1972. Erythropoietin effects on fetal mouse erythroid cells. II. Nucleic acid synthesis and the erythropoietin-sensitive cell. *J. Biol. Chem.* 247:731–35
64. Gallagher, N. I., Lange, R. D. 1962. Response to erythropoietin. *Proc. Soc. Exp. Biol. Med.* 110:422–26
65. Paul, J., Conkie, D., Burgos, H. 1973. Effects of erythropoietin on cell populations and macromolecular

synthesis in foetal mouse erythroid cells. *J. Embryol. Exp. Morphol.* 29:453–72

66. Gross, M., Goldwasser, E. 1971. On the mechanism of erythropoietin-induced differentiation. IX. Induced synthesis of 9S ribonucleic acid and of hemoglobin. *J. Biol. Chem.* 246:2480–86
67. Krantz, S. B., Goldwasser, E. 1965. On the mechanism of erythropoietin-induced differentiation. II. The effect on RNA synthesis. *Biochim. Biophys. Acta* 103:325–32
68. Gross, M., Goldwasser, E. 1969. On the mechanism of erythropoietin-induced differentiation. V. Characterization of the ribonucleic acid formed as a result of erythropoietin action. *Biochemistry* 8:1795–1805
69. Nicol, A. G., Conkie, D., Langon, W. G., Drewienkiewicz, C. E., Williamson, R., Paul, J. 1972. Characteristics of erythropoietin-induced RNA from fetal mouse liver erythropoietic cell cultures and the effects of 5-fluorodeoxyuridine. *Biochim. Biophys. Acta* 277:342–53
70. Paul, J., Hunter, J. A. 1969. Synthesis of macromolecules during induction of haemoglobin synthesis by erythropoietin. *J. Mol. Biol.* 42:31–41
71. Lockard, R. E., Lingrel, J. B. 1969. The synthesis of mouse hemoglobin β-chains in a rabbit reticulocyte cell-free system programmed with mouse reticulocyte 9S RNA. *Biochem. Biophys. Res. Commun.* 37:204–12
72. Ramirez, F., Gambino, R., Maniatis, G. M., Rifkind, R. A., Marks, P. A., Bank, A. 1975. Changes in globin messenger RNA content during erythroid cell differentiation. *J. Biol. Chem.* 250:6054–58
73. Terada, M., Cantor, L., Metafora, S., Rifkind, R. A., Bank, A., Marks, P. 1972. Globin messenger RNA activity in erythroid precursor cells and the effect of erythropoietin. *Proc. Natl. Acad. Sci. USA* 69:3575–79
74. Spivak, J. L. 1976. Effects of erythropoietin on chromosomal protein synthesis. *Blood* 47:581–92
75. Piantadosi, C. A., Dickerman, H. W., Spivak, J. L. 1976. Sequential activation of splenic nuclear RNA polymerases by erythropoietin. *J. Clin. Invest.* 57:20–26
76. Dukes, P. P. 1968. *In vitro* studies on DNA synthesis of bone marrow cells stimulated by erythropoietin. *Ann. NY Acad. Sci.* 149:437–44
77. Gross, M., Goldwasser, E. 1970. On the mechanism of erythropoietin-induced differentiation. VII. The relationship between stimulated deoxyribonucleic acid synthesis and ribonucleic acid synthesis. *J. Biol. Chem.* 245:1632–36
78. Gallien-Lartique, O., Goldwasser, E. 1965. On the mechanism of erythropoietin induced differentiation. I. The effects of specific inhibitors on hemoglobin synthesis. *Biochim. Biophys. Acta* 103:319–24
79. Krantz, S. B., Goldwasser, E. 1965. On the mechanism of erythropoietin-induced differentiation. IV. Some characteristics of erythropoietin action on hemoglobin synthesis in marrow cell culture. *Biochim. Biophys. Acta* 108:455–62
80. Chang, S. C.-S., Sikkema, D., Goldwasser, E. 1974. Evidence for an erythopoietin receptor protein on rat bone marrow cells. *Biochem. Biophys. Res. Commun.* 57:399–405
81. Brown, J. E., Adamson, J. W. 1974. *Blood* 44(6):913 (Abstr.)
82. Graber, S. E., Carrillo, M., Krantz, S. B. 1974. Lack of effect of erythropoietin on cyclic adenosine $3',5'$-monophosphate levels in rat fetal liver cells. *J. Lab. Clin. Med.* 83:288–95
83. Brown, J. E., Adamson, J. W. 1977. Studies of the influence of cyclic nucleotides on *in vitro* hemoglobin synthesis. *Br. J. Haematol.* 35:193–208
84. Brown, J. E., Adamson, J. W. 1977. Modulation of in vitro erythropoiesis. The influence of β-adrenergic agonists on erythroid colony formation. *J. Clin. Invest.* 60:70–77
85. Hadden, J. W., Hadden, E. M., Haddox, M. K., Goldberg, N. D. 1972. Guanosine $3':5'$-cyclic monophosphate: A possible intracellular mediator of mitogenic influences in lymphocytes. *Proc. Natl. Acad. Sci. USA* 69:3024–27
86. Seifert, W. E., Rudland, P. S. 1974. Possible involvement of cyclic GMP in growth control of cultured mouse cells. *Nature* 248:138–40
87. Rodgers, G. M., Fisher, J. W., George, W. J. 1976. Elevated cyclic GMP concentrations in rabbit bone

marrow culture and mouse spleen following erythropoietic stimulation. *Biochem. Biophys. Res. Commun.* 70:287–94

88. Graber, S. E., Bomboy, J. D. Jr., Salmon, W. D. Jr., Krantz, S. B. 1977. Effect of erythropoietin preparations on cyclic AMP and cyclic GMP levels in rat fetal liver cell cultures. *J. Lab. Clin. Med.* 90:162–70
89. Chang, C.-S., Goldwasser, E. 1973. On the mechanism of erythropoietin-induced differentiation. XII. A cytoplasmic protein mediating induced nuclear RNA synthesis. *Dev. Biol.* 34:246–54
90. Ward, H. P., Kurnick, J. E., Pisarczyk, M. J. 1971. Serum level of erythropoietin in anemias associated with chronic infection, malignancy, and primary hematopoietic disease. *J. Clin. Invest.* 50:332–35
91. Schooley, J. C., Mahlmann, L. J. 1971. Inhibition of the biologic activity of erythropoietin by neuraminidase *in vivo*. *J. Lab. Clin. Med.* 78:765–70
92. Reissmann, K. 1964. Protein metabolism and erythropoiesis. I. The anemia of protein deprivation. *Blood* 23:137–45
93. Krantz, S. B., Jacobson, L. O. 1970. See Ref. 3, pp. 151–64
94. Fried, W., Jonasson, O., Lang, G., Schwartz, F. 1973. The hematologic effect of androgen in uremic patients. *Ann. Intern. Med.* 79:823–27
95. Larsen, O. A., Josephson, P., Lassen, N. A. 1963. Nefrogen anemia behändlet med erythropoietin. *Ugeskr. Laeg.* 125:435–38
96. Gurney, C. W. 1966. Plasma-erythropoietin bei nierenkrankheiten. In *Symposium der gesellschaft fur nephrologie, IV. Aktuelle Probleme der nephrologie*, ed. F. Kruck, pp. 310–32. Wien: Springer-Verlag
97. Lichtman, M. A., Miller, D. R. 1970. Erythrocyte glycolosis, 2,3-diphosphoglycerate, and adenosine triphosphate concentration in uremic subjects: relationship to extracellular phosphate concentration. *J. Lab. Clin. Med.* 76:267–79
98. Wallner, S. F., Kurnick, J. E., Ward, H. P., Vautrin, R., Alfrey, A. C. 1976. The anemia of chronic renal failure and chronic diseases: *In vitro* studies of erythropoiesis. *Blood* 47:561–69
99. Van Dyke, D. 1967. *Erythropoietin therapy in the renoprival patient.* US Atomic Energy Commission, Univ. Calif. Radiation Lab. (Berkeley), UCRL-17481:127–32
100. Stamatoyannopoulos, G., Bellingham, A. J., Lenfant, C., Finch, C. A. 1971. Abnormal hemoglobins with high and low oxygen affinity. *Ann. Rev. Med.* 22:221–34
101. Cartier, P., Labie, D., Leroux, J. P. 1972. Deficit familial en diphosphoglycerate-mutase: etude hematologique et biochimique. *Nouv. Rev. Fr. Hematol.* 12:269–87
102. Adamson, J. W. 1968. The erythropoietin hematocrit relationship in normal and polycythemic man: implications of marrow regulation. *Blood* 32:597–609
103. Krantz, S. B., Jacobson, L. O. 1970. See Ref. 3, pp. 184–95
104. Adamson, J. W., Stamatoyannopoulus, G., Kontras, S., Lascari, S., Detter, N. 1973. Recessive familial erythropoiesis: aspects of marrow regulation in two families. *Blood* 41:641–52
105. Krantz, S. B. 1968. Response of polycythemia vera marrow to erythropoietin *in vitro*. *J. Lab. Clin. Med.* 71:999–1012
106. Moriyama, Y. 1969. Studies on regulation of erythropoiesis in polycythemia vera. *Jpn. J. Clin. Invest.* 10:701–11
107. Prchal, J. F., Adamson, J. W., Murphy, S., Steinmann, L., Fialkow, P. J. 1976. *Clin. Res.* 24(3):442 (Abstr.)
108. Golde, D. W., Bersch, N., Cline, M. J. 1977. Polycythemia vera: Hormonal modulation of erythropoiesis *in vitro*. *Blood* 49:399–405
109. Zanjani, E. D., Lutton, J. D., Hoffman, R., Wasserman, L. R. 1977. Erythroid colony formation by polycythemia vera bone marrow *in vitro*. *J. Clin. Invest.* 59:841–48

Ann. Rev. Med. 1978. 29:67–72
Copyright © 1978 by Annual Reviews Inc. All rights reserved

RECURRENT GLOMERULONEPHRITIS AFTER RENAL TRANSPLANTATION

◆7266

Jean Hamburger, M.D., Jean Crosnier, M.D., and Laure-Hélène Noël, M.D.
Département de Néphrologie, Hôpital Necker, Paris, France

The first mention that glomerulonephritis (GN) could recur on a transplanted kidney was made by Glassock et al (1) in a series of renal transplantations between identical twins. The absence of rejection phenomena in these isografts was strong evidence that the glomerular disease of the transplant was a recurrence of the patient's initial GN. In Glassock's series, 11 out of 18 patients with renal failure resulting from GN developed clinical evidence of a similar disease in the transplant. Seven of these grafts were destroyed by the recurrent GN. Such a high incidence was not found, however, in subsequent series and in allografts, raising the possibility that immunosuppressive drugs may exert some protective action.

In 1964 our group (2) described another type of glomerular lesion in renal allografts when the patient's own kidney had been destroyed by a dissimilar disease such as a congenital or familial abnormality. We suggested that this could be a form of rejection. This "rejection glomerular disease" was confirmed by Porter (3) and many others.

In addition to these two main causes of glomerular damage in renal grafts, other possibilities include *de novo* GN following for example a hepatitis B (4), glomerular alterations provoked by the administration of antilymphocyte serum (5), and transplantation of a kidney already altered by latent GN (6). McPhaul et al (7) and Cameron & Turner (8) have recently reviewed these various possibilities. This paper focuses mainly on recurrent glomerulonephritis and rejection glomerular lesions, not only because they represent a large majority of transplant glomerular lesions, but also because they afford new insights into the pathogenesis of GN (9).

Our study is based on reports from the literature and on 166 personal cases (Table 1) from a series of approximately 500 patients.

0066-4219/78/0401-0067$01.00

TRANSMISSION OF THE PATIENT'S ORIGINAL GN TO THE GRAFT

The main types of chronic GN subject to recurrence in the graft are the nephrotic syndrome with focal glomerular sclerosis, membranous GN, membrano-proliferative GN, rapidly progressive GN, and Immunoglobulin A (IgA) mesangial deposit nephropathy.

In our series, recurrence has been observed in 17.5% of patients whose own kidneys had been destroyed by GN. The subsequent course is, however, far from being severe in all these cases.

Recurrence of the Nephrotic Syndrome With Focal and Segmental Glomerulosclerosis

Recurrence was observed in 4 out of 14 allografted patients in our series. Similar results have been reported by Hoyer et al (10), Habib (11), Brown et al, and many others [see references in (8)]. Special features usually characterize recurrence in these cases. Massive proteinuria, often seen in the hours immediately following transplantation, may be followed rapidly by a nephrotic syndrome, which is often uninfluenced by corticosteroids. Biopsy shows only minimal changes during this initial stage. Some months later segmental and focal sclerosis of the graft glomeruli may appear, with focal deposits of IgM, C_3, and rarely IgG. Loss of the graft because of recurrence is far from constant. In two patients receiving second transplants, immediate recurrence was seen in the second as well as the first graft. Recurrence in this form of GN seems the more frequent and severe since the course of the primary disease toward renal failure was shorter.

Table 1 Incidence of recurrent glomerulonephritis (GN) in the series of Necker Hospital, Paris, France

Type of GN	No. of patients	No. of cases with recurrence
Focal and segmental glomerulosclerosis	14	4
Membranous GN	2	0
Membrano-proliferative GN with subendothelial deposits	23	2
Membrano-proliferative GN with isolated C_3 deposits	4	3
Membrano-proliferative GN, lobular form	13	2
Dense deposits disease	11	8
Rapidly progressive GN	1	0
IgA mesangial deposits	20	9
Schönlein-Henoch purpura	6	1
Unclassified GN	72	
Total	166	29

These observations may be important for the understanding of the still mysterious problem of "lipoid nephrosis." They suggest that a permanent humoral abnormality altering glomerular basement-membrane permeability to macromolecules may be present, at least in severe cases of nephrotic syndrome inducing rapid renal failure.

Recurrence of Membranous Glomerulonephritis

Recurrence very rarely has been reported in GN with extramembranous deposits. According to Cameron & Turner, evidence of recurrence is firmly established in only two published cases (8). It may seem surprising that recurrence appears rare in a type of GN considered a result of circulating-immune complexes whose formation may continue for a long time. But it must be pointed out that membranous GN often has a benign or slow course so that a limited number of cases require transplantation.

Recurrence of Membrano-Proliferative Glomerulonephritis

In the common form of membrano-proliferative GN with subendothelial deposits, the study of recurrence is difficult because clinical and pathological findings are similar to those of the "rejection glomerular disease" described below. In our series, only 2 of 23 cases were classified as certain recurrences. More frequent, or perhaps more easily demonstrated, was recurrence in the three special forms which follow.

Membrano-proliferative GN with C_3 deposits (no immunoglobulin deposit) is prone to recur with the same immunofluorescent pattern and persistently low level of C_3 in the serum. In our cases, clinical manifestations were limited and the course was benign.

In the "lobular" type of proliferative nephritis, an identical lesion can appear on the allograft. This was observed in 2 of 13 transplanted patients. One of these cases was studied with serial biopsies. There first appeared mesangial deposits of IgG and C_3, then subepithelial deposits, and only thereafter endocapillary proliferation and formation of lobular nodules (12). This observation of the development of a nephritis in a normal kidney is of course important for the understanding of spontaneous GN.

Finally, the "dense deposits disease" described by Galle & Berger in 1963, associating endocapillary proliferation with intramembranous, electron-dense deposits of a still unknown substance, recurs in a large number of cases: 8 patients of 11 in our series and at least 18 others in the literature (13, 14). The recurrent lesion can remain latent but nonetheless, these cases are evidence that the lesion depends upon some systemic factors. The serum C_3 level, which is almost constantly low in this disease, may either remain low or return to normal after transplantation, even when biopsy shows development of the membrane lesion. In addition, C_3 nephritic factor (NF) serum activity, commonly detected in this type of GN, often disappears soon after transplantation (15).

Recurrence of Rapidly Progressive (Malignant) Glomerulonephritis

Most studies concern cases of crescentic GN with linear IgG deposits along the glomerular capillary wall, with or without anti–glomerular basement membrane (GBM) antibodies detectable in the serum or in kidney eluates (15–19). In such cases, a similar linear IgG fixation has been reported in the transplanted kidney, with or without crescent formation. This has been seen even in patients with no circulating anti-GBM antibodies, although they were found in transplant eluates (20). These patients may show clinical manifestations, but several cases of recurrent linear deposit without kidney damage have been reported (18). Dixon & Wilson (17, 19) report that of transplantation in 15 patients with anti-GBM disease (excluding Goodpasture's syndrome), 6 showed clinical or immunofluorescent evidence of recurrence and graft failure ensued in at least 3. Anti-GBM antibodies may fix on the graft within minutes after transplantation (16) but when no linear IgG fixation is found on the first biopsy, it never appears subsequently.

Goodpasture's syndrome may also be mentioned in this regard. The 1971 Human Renal Transplant Registry reported 12 patients with this syndrome having undergone transplantation; in 8 of them the graft was unaltered after 3 months to 5 years.

Recurrence of IgA Mesangial Deposits

Also seen to recur in the transplant is the special form of glomerulonephritis described by Berger in 1969, associating diffuse mesangial deposits of IgA and IgG with usually segmental and focal glomerular lesions. Early IgA mesangial deposits were seen in 9 of 20 patients in our series (21) and in 2 patients reported by Mathew et al (22). The diffuse IgA deposits often remain latent with excellent allograft function for as long as 8 years after transplantion.

Interestingly enough, some types of secondary GN associated with a systemic disease may have a low incidence of recurrence, if any. We have already mentioned Goodpasture's syndrome. The 1975 Renal Transplant Registry reports 56 patients with systemic lupus receiving renal transplants and never showing recurrence either on biopsy specimens or autopsy material. Here again, the concept that circulating immune complexes explain most cases of GN is not supported by observations on transplanted kidneys. In the GN of Schönlein-Henoch purpura, only three cases of recurrence have been reported [see references in (23).]. Of six patients in our series, only one showed histological evidence of recurrence.

GLOMERULAR LESIONS PROBABLY RELATED TO RECIPIENT IMMUNE RESPONSE TO THE GRAFT: THE "REJECTION GLOMERULAR DISEASE"

Since our early description of "rejection glomerulonephritis" (2, 24), such lesions have been described by many authors (3, 25–28). When minimal lesions visible only on electron microscopy are included, no less than 50% of transplanted patients show this kind of glomerular alteration. The commonest pathological picture includes deposits of IgM, IgG, and C_3; mesangial cell proliferation frequently moderate and focal; and membranous changes. Clinical manifestations are proteinuria, sometimes the nephrotic syndrome, a microscopic hematuria, and a possible course towards renal insufficiency. Many of these cases have a benign course. One of our patients having these lesions is still in perfect condition more than ten years after transplantation, with normal plasma creatinine and urea and normal function tests.

The significance of these lesions is still unknown. Because of the immunoglobulin deposits, it has been suggested that circulating immune complexes could be induced by the antigraft response of the host. However, the relation between immune deposits and glomerular lesions has not been completely elucidated in these cases. On the one hand, immunoglobulin and complement deposits may be found in transplanted kidneys with no glomerular lesions, inducing no glomerular cell proliferation at any time. Conversely, little or no immunoglobulin or complement deposition may be found, despite glomerular cell proliferation and other alterations in the glomerulus. It is true that deposits may have existed for a short time between two biopsies. But these observations cannot be ignored when discussing the mechanism of spontaneous GN: Glomerular lesions of transplanted kidneys may almost be considered accidental human-experimental GN (9).

Literature Cited

1. Glassock, R. J., Feldman, D., Reynolds, E. S., Dammin, G. J., Merrill, J. P. 1968. Human renal isografts: a clinical and pathologic analysis. *Medicine* 47:411–54
2. Hamburger, J., Crosnier, J., Dormont, J. 1964. Observations in patients with a well tolerated homotransplanted kidney. *Ann. NY Acad. Sci.* 120:558–77
3. Porter, K. A., Rendall, J. M., Stolinski, C., Terasaki, P. I., Marchioro, T. L., Starzl, T. E. 1966. Light and microscopic study of biopsies from 33 human renal allografts and an isograft 1¾ to 2 years after transplantation. *Ann. NY Acad. Sci.* 129:615–36
4. Sonnabend, W., Kistler, G. S., Thiel, G., Zollinger, H. U., Bianchi, L., Enderlin, F., Gudat, F., Krech, U. 1974. Chronische Herdglomerulitis und chronisch-persistierende Hepatitis nach Nierentransplantation: Nachweis von Hepatitis B und Epstein-Barr-virus Antigen. *Schweiz. Med. Wochenschr.* 104:1205–17
5. Busch, G. J., Birtch, A. G., Luke, P. Jr., Kobayashi, K., Galvanek, E. G., Carpenter, C. B. 1971. Human renal allografts. Glomerular deposits of horse immunoglobulin G and nephritis following administration of antilymphocyte globulin. *Hum. Pathol.* 2:299–308
6. Brutel de la Riviere, G., van de Putte, L. B. A. 1976. Pre-existing glomerulonephritis in allografted kidneys. *Arch. Pathol.* 100:196–98
7. McPhaul, J. J., Lordon, R. E., Thompson, A. L., Mullins, J. D. 1976. Nephritogenic immunopathologic mechanisms and human renal transplants. The problem of recurrent glomerulonephritis. *Kidney Int.* 10:135–38
8. Cameron, J. S., Turner, D. R. 1977. Recurrent glomerulonephritis in allografted kidneys. *Clin. Nephrol.* 7:47–54
9. Hamburger, J., Berger, J., Hinglais, N., Descamps, B. 1973. New insights into the pathogenesis of glomerulonephritis afforded by the study of renal allografts. *Clin. Nephrol.* 1:3–7
10. Hoyer, J. R., Raij, L., Vernier, R. L., Simmons, R. L., Najarian, J. S., Michael, A. F. 1972. Re-

currence of idiopathic nephrotic syndrome after renal transplantation. *Lancet* 2:343–48

11. Habib, R. 1973. Focal glomerular sclerosis. *Kidney Int.* 4:355–61
12. Schürch, W., Leski, M., Hinglais, N. 1972. Evolution of recurrent lobular glomerulonephritis in a human kidney allotransplant. Combined light, immunofluorescence and electron microscopic studies of serial biopsies. *Virchows Arch. A* 355:66–84
13. Galle, P., Hinglais, N., Crosnier, J. 1971. Recurrence of an original glomerular lesion in three renal allografts. *Transplant. Proc.* 3:368–70
14. Turner, D. R., Cameron, J. S., Bewick, M., Sharpstone, P., Melcher, D., Ogg, C. S., Evans, D. I., Trafford, A. J. P., Leibowitz, S. 1976. Transplantation in mesangiocapillary glomerulonephritis with intramembranous dense "deposits." Recurrence of disease. *Kidney Int.* 9:439–48
15. Droz, D., Zanetti, M., Noël, L. H., Leibowitch, J. 1977. Dense deposits disease. *Nephron.* 19:1–11
16. Lerner, R. A., Glassock, R. J., Dixon, F. J. 1967. The role of antiglomerular basement membrane antibody in the pathogenesis of human glomerulonephritis. *J. Exp. Med.* 126:989–1004
17. Dixon, F. J., McPhaul, J. J., Lerner, R. A. 1969. The contribution of kidney transplantation to the study of glomerulonephritis. The recurrence of glomerulonephritis in renal transplants. *Transplant. Proc.* 1:191–96
18. Beleil, O. M., Coburn, J. W., Shinaberger, J. H., Glassock, R. J. 1973. Recurrent glomerulonephritis due to anti-glomerular basement membrane antibodies in two successive allografts. *Clin. Nephrol.* 1:377–80
19. Wilson, C. B., Dixon, F. J. 1973. Antiglomerular basement membrane antibody induced glomerulonephritis. *Kidney Int.* 3:74–89
20. Couser, W. G., Wallace, A., Monaco, A. P., Lewis, E. J. 1973. Successful renal transplantation in patients with circulating antibody to glomerular basement membrane: Report of two cases. *Clin. Nephrol.* 1:381–88
21. Berger, J., Yaneva, H., Nabarra, B., Barbanel, C. 1975. Recurrence of mesangial deposition of IgA after renal transplantation. *Kidney Int.* 7:232–41
22. Mathew, T. H., Mathews, D. C., Hobbs, J. B., Kincaid-Smith, P. 1975. Glomerular lesions after renal transplantation. *Am. J. Med.* 59:177–90
23. Levy, M., Broyer, M., Arsan, A., Levy-Bentolila, D., Habib, R. 1976. Anaphylactoid purpura nephitis in childhood: Natural history and immunopathology. *Adv. Nephrol. Necker Hosp.* 6:183–228
24. Hamburger, J., Vaysse, J., Crosnier, J., Auvert, J., Dormont, J. 1963. Résultats de quatorze transplantations rénales chez l'homme. *Bull. Mem. Soc. Méd. Hôp. Paris* 114:225–30
25. Porter, K. A., Dossetor, J. B., Marchioro, T. L., Peart, W. S., Rendall, J. M., Starzl, T. C., Terasaki, P. I. 1967. Human renal transplants. I. Glomerular changes. *Lab. Invest.* 16:153–81
26. Shimamura, T., Gyorkey, F., Morgen, R. O., Debakey, M. E. 1966. The fine structural observations in human kidney homografts. *Invest. Urol.* 3:590–98
27. Lindquist, R. R., Guttman, R. D., Merrill, J. P., Dammin, G. J. 1968. Human renal allografts. *Am. J. Pathol.* 53:851–81
28. Olsen, S., Bolman, S., Petersen, V. P. 1974. Ultrastructure of the glomerular basement membrane in long term renal allografts with transplant glomerular disease. *Lab. Invest.* 30:176–89

Ann. Rev. Med. 1978. 29:73–79
Copyright © 1978 by Annual Reviews Inc. All rights reserved

NEURAL ACTIONS OF DIGITALIS ◆7267

Richard A. Gillis, Ph.D. and John A. Quest, Ph.D.
Department of Pharmacology, Georgetown University Schools of Medicine and Dentistry, Washington DC 20007

INTRODUCTION

Neural actions of digitalis drugs have been well documented but their role in the cardiovascular effects of this group of agents is a subject of considerable controversy. Information concerning neural actions of these drugs, as related to their cardiovascular effects, can be grouped into the following three categories: conduction, automaticity, and refractory period; ventricular arrhythmias; and inotropic and hemodynamic effects. The first are examples where firm evidence from both animal and human studies indicates a major role for the nervous system in determining the response produced by digitalis drugs. The second is an example where impressive evidence has been obtained from some but not all animal studies indicating an important role for the nervous system in the toxic response of these agents. The third are examples where present data suggest a role for the nervous system in the well-known actions of digitalis drugs on the heart and vasculature, but more experimentation is needed to confirm present findings and to further elucidate the extent to which the nervous system participates in these responses.

CONDUCTION, AUTOMATICITY, AND REFRACTORY PERIOD

Experimental and clinical studies indicate that the entire effect of digitalis on atrioventricular conduction is mediated through the nervous system. In animals, slowing in conduction induced by digitalis can be completely prevented by either pharmacological blockade (1) or surgical interruption (2) of autonomic influences on the heart. Likewise, in humans, slowing in conduction is completely prevented by surgical denervation of the heart, as has been demonstrated in studies performed in patients with transplanted hearts (3, 4). The neural mechanisms involved in the conduction effect of digitalis are enhancement of cardiac vagal activity and inhibition of cardiac sympathetic activity (4).

Likewise, digitalis-induced slowing in sinus rate is mediated entirely through the nervous system. This has been demonstrated in animals (5, 6), as well as in

0066-4219/78/0401-0073$01.00

humans (3, 4). When the parasympathetic and sympathetic neural influences on the sinus node have been eliminated, no significant decrease in sinus rate has been observed with either therapeutic doses or doses bordering on the toxic range. The near-toxic dose produces an increase in sinus rate (5, 6), which appears to be also mediated by the nervous system. That is, the increase in sinus rate is related to digitalis-induced release of norepinephrine from cardiac stores (6).

The nervous system also plays a prominent role in the atrial refractory-period effects of digitalis. With the nervous system intact, a shortening in the atrial refractory period occurs. Animal studies demonstrate that this shortening is due primarily to augmentation of vagal activity on the heart (7). Studies in humans support these findings since no shortening of the atrial refractory period was observed when digoxin was administered to subjects with transplanted hearts (3). In the absence of neural influences, digitalis induces an increase in atrial refractory period (3), an action apparently due to a direct effect on atrial tissue (7).

VENTRICULAR ARRHYTHMIAS

There is considerable evidence to indicate that digitalis in toxic doses can enhance the arrhythmogenic influence of the sympathetic nervous system. In toxic doses these drugs have been shown to augment neural discharge in both cardiac pre- and postganglionic sympathetic nerves (8). In addition, enhanced release of catecholamines occurs (9). These effects may reflect a stimulatory effect of digitalis either in the brain (8), spinal cord (9), or peripheral ganglia (10). Furthermore, Roberts and colleagues (11) have obtained evidence to indicate that digitalis produces nonuniform changes in postganglionic cardiac sympathetic-nerve filaments. This nonuniform activity may in turn result in nonuniform alterations in excitability and conduction in the myocardium. Finally, toxic doses of digitalis enhance the sensitivity of β-adrenergic receptors in the ventricular myocardium (12).

That the above changes in sympathetic function induced by digitalis result in cardiac arrhythmias is indicated by the following: (*a*) a temporal correlation exists between digitalis-induced alterations in sympathetic activity and the development of cardiac arrhythmias; (*b*) abolition of digitalis-induced sympathetic hyperactivity is associated with a simultaneous conversion of a ventricular arrhythmia to sinus rhythm; (*c*) animals with a depressed or absent sympathetic nervous system are resistant to the arrhythmogenic effects of digitalis; (*d*) ventricular arrhythmias can be produced by subthreshold levels of electrical stimulation applied to the hypothalamus during systemic administration of subarrhythmic doses of digitalis; and (*e*) animals with depressed sympathetic nervous system function exhibit a lower incidence of ventricular fibrillation (13, 14).

There have been few studies on humans aimed at determining the role of the sympathetic nervous system in the arrhythmias produced by digitalis. One approach has been to examine the ability of β-adrenergic blocking drugs to influence digitalis-induced cardiac arrhythmias (15). From the numerous studies performed, it is clear that β-adrenergic blocking doses of these compounds can counteract both atrial and ventricular arrhythmias induced by digitalis. This is also true with β-adrenergic blocking drugs that possess relatively little direct nonspecific-membrane effects (15). These data from human studies support the concept that digitalis-induced alterations in sympathetic function are important components of the arrhythmogenic effects of digitalis.

INOTROPIC AND HEMODYNAMIC CHANGES

A question that has been raised and discussed in numerous articles is whether or not the positive inotropic effect of digitalis is mediated by the sympathetic nervous system. The conclusion reached is that it is not (16). This conclusion is based on either in vivo studies performed in animals with cardiovascular reflexes intact, or in vitro studies using atrial, ventricular, or papillary muscle preparations. However, in one in vivo study using animals with denervated cardiovascular reflexes, evidence suggested that an action of digitalis on forebrain structures results in sympathetic activation and a concomitant increase in contractile force (17).

A second question, raised initially by Daggett & Weisfeldt (18), is whether neural effects of digitalis drugs are capable of obscuring the positive inotropic and/or the cardiac output effects of these agents in normal animals. This question has not been resolved. Studies using decerebrate and also anesthetized animals indicate that the reflex actions of digitalis mask the positive inotropic effect of the drug (19, 20). This has not been confirmed in a study using unanesthetized dogs (21), but the study can be criticized on two points. First, the cardiovascular reflexes were not totally removed as vagal-afferent neural fibers were left intact, and second, the dose of digitalis employed was not as large as that used in the anesthetized animal studies.

There is also evidence that the vascular effects of digitalis are mediated through the nervous system, specifically through either activation or inhibition of sympathetic neural effects on arterial and venous smooth muscle. Activation has been demonstrated for the pressor response observed with ouabain in cats with denervated reflexes (22); coronary constriction observed with acetylstrophanthidin and digoxin in dogs (23); skeletal muscle vascular-bed constriction observed with acetylstrophanthidin, ouabain, and dihydro-ouabain in dogs (24, 25); and contraction of the cutaneous and saphenous veins observed with acetylstrophanthidin and digoxin in dogs (26). Inhibition of the sympathetic nervous system and consequent vasodilation has been demonstrated for arterial and venous beds of experimental animals (19, 27) and humans (28, 29). Mason and colleagues (28) have demonstrated that arterial dilation and venodilation occur when digitalis is used in humans with congestive heart failure. As expected, arterial dilation resulted in a decrease in total peripheral resistance which would contribute to the beneficial response on cardiac output seen with digitalis. Mason and colleagues speculated that the venodilation would result in increased venous return in patients with heart failure.

CLINICAL IMPLICATIONS AND NEW DIRECTIONS FOR RESEARCH

Most of the important clinical effects of digitalis, as well as the role of the nervous system in each effect, and the mechanism(s) underlying each neural action are tabulated in Table 1. Digitalis-induced neural effects are clearly important in using these agents for treating atrial arrhythmias. Their effect of slowing ventricular rate in patients with either atrial flutter or fibrillation is due to neurally-induced changes in atrioventricular conduction. Part of the slowing in ventricular rate that occurs in some patients with atrial flutter is due to atrial fibrillation developing and enhancing concealed conduction. This effect is also mediated by the nervous system, specifically digitalis-induced parasympathomimetic effects that result in shortening of atrial muscle-cell refractory period. Interruption of paroxysmal atrial tachycardia is also primarily due to parasympathetically mediated effects on conduction pathways in the atrio-

Table 1 Summary of neuro-cardiovascular effects of digitalis

Cardiovascular effect	Neural mechanism(s)	Site of Neural action	Effect that is neurally medicated (% approximate)	Therapeutic Application
1. Decrease in the speed of atrioventricular conduction	Activation of parasympathetic and inhibition of sympathetic nervous system	Peripheral reflexogenic sites; postsynaptic cholinergic cardiac receptors	100	Slowing of ventricular rate in patients with atrial flutter and fibrillation; treatment of paroxysmal atrial tachycardia
2. Decrease in refractory period of atrial muscle cells	Activation of parasympathetic nervous system	Same as above	100	May be responsible for converting atrial flutter to fibrillation
3. Decrease in sinus rate	Same as in 1	Same as above	100	May contribute to slowing in sinus rate of patients with congestive heart failure
4. Antagonism of ventricular arrhythmias	Same as in 1	Same as above	Not determined	See Lown et al (30)
5. Induction of ventricular tachyarrhythmias	Activation of sympathetic nervous system	CNS, ganglia, postsynaptic adrenergic cardiac receptors	Not determined	Employing drugs that reduce sympathetic outflow to treat digitalis-induced ventricular arrhythmias (31)
6. Decrease in arterial and venous tone	Inhibition of sympathetic nervous system	Peripheral reflexogenic sites	100	Treating congestive heart failure; possibly treating hypertension

ventricular node (32), though it may also be due to suppression of pacemaker activity of an abnormal atrial focus just as sinus node suppression occurs (Table 1). Recent data indicate that digitalis has an antiarrhythmic effect in patients with ventricular arrhythmias (30, 33). Lown and colleagues (30) attribute this antiarrhythmic effect to (*a*) enhanced cardiac vagal activity which in turn antagonizes norepinephrine release from cardiac nerves, and (*b*) an antagonistic action of the two neurotransmitters at the ventricular receptor level. Another possible mechanism may be withdrawal of sympathetic tone induced by an action of digitalis on the reflexes (18, 34, 35).

Digitalis-induced neural effects may also be important in the treatment of congestive heart failure. Mason and colleagues (28) concluded that neural effects may be involved in the reduced resistance to ventricular ejection that occurs when digitalis is administered to patients with congestive failure. They conclude that the neural effect occurs indirectly through the baroreceptors, and results from the improved cardiac output induced by the direct positive inotropic effect of these drugs. Based on studies in animals (36) and patients (37) which indicate that digitalis sensitizes baroreceptors, reduced resistance may be due to a direct effect on the sensory elements of the reflexogenic areas. These same investigators also postulate that digitalis induces an increase in venous return in patients with congestive failure. The increase results from a "shift of blood in the venous reservoir from the portal to the systemic bed" (28). The shift of blood is possibly due to the "release of sympathetically-induced venoconstriction which is more intense in hepatic veins" (28). This release is again postulated to result indirectly (via baroreceptors) from the positive inotropic effect of the drug. It is possible that the direct effect of digitalis on baroreceptors is primarily responsible for this effect. In fact, if a direct neural effect is responsible, it should occur prior to the inotropic effect and result in a decrease in venous return (27) which would be beneficial to the heart in failure.

Another potential but as yet unstudied effect that might be mediated by the reflexogenic effects of digitalis is diuresis. Sensory stretch receptors located in the left atrium exert control over antidiuretic hormone secretion (38) and therefore influence urine output. Activation of these receptors decreases secretion of antidiuretic hormone and increases sodium and water excretion. Digitalis administered to patients in congestive failure might sensitize these receptors and result in a diuretic response. These reflex effects of digitalis assume greater importance in view of the fact that stretch receptors appear to be desensitized in these patients (39). Administration of digitalis may shift the sensitivity of these receptors toward the normal.

The role of the nervous system in toxic-vasoconstrictor responses in humans has not been fully assessed, but there is a great need to do so. Vasoconstriction, which occurs primarily when digitalis is administered to individuals not in heart failure, is detrimental and can lead to acute hypertension (40), heart failure (41), mesenteric infarction (42), and possibly angina (43). If sympathetic neural activation contributes to the vasoconstrictor response, then pretreatment or treatment with α-adrenergic blocking agents may be beneficial to patients exhibiting, or likely to exhibit, this effect. Indeed, it has been recently demonstrated that the vasoconstrictor effect of dihydro-ouabain, unlike that of other digitalis preparations, can be almost completely blocked by pretreatment with an α-adrenergic receptor blocking agent (25).

Literature Cited

1. Kim, Y. I., Noble, R. J., Zipes, D. P. 1975. Dissociation of the inotropic effect of digitalis from its effect on atrioventricular conduction. *Am. J. Cardiol.* 36:459–67
2. Schaal, S. F., Sugimoto, T., Wallace, A. G., Sealy, W. C. 1968. Effects of digitalis on the functional refractory period of the A-V node: Studies in awake dogs with and without cardiac denervation. *Cardiovasc. Res.* 4:356–59
3. Leachman, R. D., Cokkinos, D. V. P., Cabrera, R., Leatherman, L. L., Rochelle, D. G. 1971. Response of the transplanted, denervated human heart to cardiovascular drugs. *Am. J. Cardiol.* 27:272–76
4. Goodman, D. J., Rossen, R. M., Cannom, D. S., Rider, A. K., Harrison, D. C. 1975. Effect of digoxin on atrioventricular conduction: Studies in patients with and without cardiac autonomic innervation. *Circulation* 51:251–56
5. Ten Eick, R. E., Hoffman, B. F. 1969. Chronotropic effect of cardiac glycosides in cats, dogs, and rabbits. *Circ. Res.* 25:365–78
6. Gillis, R. A., Quest, J. A., Thibodeaux, H., Clancy, M. M., Evans, D. E. 1975. Neural mechanisms involved in acetylstrophanthidin-induced bradycardia. *J. Pharmacol. Exp. Ther.* 193:336–45
7. Farah, A., Loomis, T. A. 1950. The action of cardiac glycosides on experimental atrial flutter. *Circulation* 2:742–48
8. Pace, D. G., Gillis, R. A. 1976. Neuroexcitatory actions of digoxin in the cat. *J. Pharmacol. Exp. Ther.* 199:583–600
9. Hashimoto, K., Kimura, T., Kubota, K. 1973. Study of the therapeutic and toxic effects of ouabain by simultaneous observations on the excised and blood-perfused sinoatrial node and papillary muscle preparations and the *in situ* heart of dogs. *J. Pharmacol. Exp. Ther.* 186:463–71
10. Weaver, L. C., Akera, T., Brody, T. M. 1976. Digoxin toxicity: primary sites of drug action on the sympathetic nervous system. *J. Pharmacol. Exp. Ther.* 197:1–9
11. Roberts, J., Kelliher, G. J., Lathers, C. M. 1976. Role of adrenergic influences in digitalis-induced ventricular arrhythmia. *Life Sci.* 18:665–78
12. Pearle, D. L., Gillis, R. A. 1974. Effect of digitalis on the response of ventricular pacemakers to sympathetic nerve stimulation. *Am. J. Cardiol.* 34:704–10
13. Gillis, R. A., Corr, P. B., Pace, D. G., Evans, D. E., DiMicco, J., Pearle, D. L. 1976. Role of the nervous system in experimentally induced arrhythmias. *Cardiology* 61:37–49
14. Erlij, D., Mendez, R. 1964. The modification of digitalis intoxication by excluding adrenergic influences on the heart. *J. Pharmacol. Exp. Ther.* 144:97–103
15. Wit, A. L., Hoffman, B. F., Rosen, M. R. 1975. Electrophysiology and pharmacology of cardiac arrhythmias IX. Cardiac electrophysiologic effects of beta adrenergic receptor stimulation and block. Part C. *Am. Heart J.* 90:795–803
16. Edmands, R. E., Greenspan, K. 1969. The inotropic effects of digitalis. A review. *J. Ind. State Med. Assoc.* 62:268–74
17. Quest, J. A., Gillis, R. A. 1975. Influences of midcollicular decerebration on the cardiovascular effects of ouabain. *Neuropharmacology* 14:763–68
18. Daggett, W. M., Weisfeldt, M. L. 1965. Influence of the sympathetic nervous system on the response of the normal heart to digitalis. *Am. J. Cardiol.* 16:394–405
19. Quest, J. A., Gillis, R. A. 1971. Carotid sinus reflex changes produced by digitalis. *J. Pharmacol. Exp. Ther.* 177:650–61
20. Reing, C. M., Thibodeaux, H., Gillis, R. A., Kot, P. A. 1973.

Influence of baroreceptor reflexes on cardiovascular responses of the dog to acetylstrophanthidin. *J. Pharmacol. Exp. Ther.* 184:641–48

21. McRitchie, R. J., Vatner, S. F. 1976. The role of arterial baroreceptors in mediating the cardiovascular response to a cardiac glycoside in conscious dogs. *Circ. Res.* 38:321–26
22. Gillis, R. A., Quest, J. A., Standaert, F. G. 1969. Depression by reflexes of the pressor and cardiotoxic responses to ouabain. *J. Pharmacol. Exp. Ther.* 170:294–302
23. Hamlin, N. P., Willerson, J. T., Garan, H., Powell, W. J. Jr. 1974. The neurogenic vasoconstrictor effect of digitalis on coronary vascular resistance. *J. Clin. Invest.* 53:288–96
24. Stark, J. J., Sanders, C. A., Powell, W. J. Jr. 1972. Neurally mediated and direct effects of acetylstrophanthidin on canine skeletal muscle vascular resistance. *Circ. Res.* 30:274–82
25. Quest, J. A., Rowles, G. S., Bhat, H. B., Gillis, R. A. 1976. Effect of dihydro-ouabain on vascular tone of the perfused canine hindlimb. *J. Pharmacol. Exp. Ther.* 199:255–61
26. Brender, D., Vanhoutte, P. M., Shepard, J. T. 1969. Potentiation of adrenergic venomotor responses in dogs by cardiac glycosides. *Circ. Res.* 25:597–606
27. Reing, C. M., Doeblin, P. L., Thibodeaux, H., Gillis, R. A., Kot, P. A. 1973. Influence of baroreceptor reflexes on the capacitance vessel's response to acetylstrophanthidin. *Eur. J. Pharmacol.* 24:341–46
28. Mason, D. T., Spann, J. F. Jr., Zelis, R. 1969. New developments in the understanding of the actions of digitalis glycosides. *Progr. Cardiovasc. Dis.* 11:443–78
29. Abarquez, R. F. Jr. 1967. Digitalis in the treatment of hypertension. A preliminary report. *Acta Med. Philipp.* 3:161–70
30. Lown, B., Graboys, T. B., Podrid, P. J., Cohen, B. H., Stockman, M. B., Gaughan, C. E. 1977. Effect of a digitalis drug on ventricular premature beats. *N. Engl. J. Med.* 296:301–6
31. Gillis, R. A., Pearle, D. L., Levitt, B. 1975. Digitalis: A neuroexcitatory drug. *Circulation* 52:739–42
32. Wu, D., Wyndham, C., Amat-Y-Leon, F., Denes, P., Dhingra, R. C., Rosen, K. M. 1975. The effects of ouabain on induction of atrioventricular nodal re-entrant paroxysmal supraventricular tachycardia. *Circulation* 52:201–7
33. Blumberg, J. M., Hayes, J. G., Stevens, M., Sullivan, G., Killip, T. 1973. Digitalis in treatment of ventricular extrasystoles in the otherwise normal heart. *Circulation* 48: Suppl. IV, p. IV–18
34. Gillis, R. A. 1969. Cardiac sympathetic nerve activity: Changes induced by ouabain and propranolol. *Science* 166:508–10
35. Evans, D. E., Gillis, R. A. 1975. Effect of ouabain and its interaction with diphenylhydantoin on cardiac arrhythmias induced by hypothalamic stimulation. *J. Pharmacol. Exp. Ther.* 195:577–86
36. Quest, J. A., Gillis, R. A. 1974. Effect of digitalis on carotid sinus baroreceptor activity. *Circ. Res.* 35:247–55
37. Thompson, M. A., Eckberg, D. L. 1976. Effects of ouabain upon the carotid baroreceptor reflex in normal man. *Circulation* 53 & 54: Suppl. II, p. II–224
38. Baum, T. 1977. Fundamental principles governing the regulation of circulatory function. In *Cardiovascular Pharmacology*, ed. M. J. Antonaccio, pp. 29–30. New York: Raven
39. Eckberg, D. L., Drabinsky, M., Braunwald, E. 1971. Defective cardiac parasympathetic control in patients with heart disease. *N. Engl. J. Med.* 285:877–81
40. Kumar, R., Yankopoulos, N. A., Abelmann, W. H. 1973. Ouabain-induced hypertension in a patient with decompensated hypertensive heart disease. *Chest* 63:105–8
41. Cohn, J. N., Tristani, F. E., Khatri, I. M. 1969. Cardiac and peripheral vascular effects of digitalis in clinical cardiogenic shock. *Am. Heart J.* 78:318–30
42. Pierce, G. E., Brockenbrough, E. C. 1970. The spectrum of mesenteric infarction. *Am. J. Surg.* 119:233–39
43. Braunwald, E., Pool, P. E. 1968. Mechanism of action of digitalis glycosides (II). *Mod. Concepts Cardiovasc. Dis.* 37:135–39

Ann. Rev. Med. 1978. 29:81–91
Copyright © 1978 by Annual Reviews Inc. All rights reserved

THE NON-HODGKIN'S LYMPHOMAS: CURRENT CONCEPTS AND MANAGEMENT

◆7268

Carol S. Portlock, M.D.
Division of Oncology, Department of Medicine, Stanford University Medical Center, Stanford, California 94305

Eli Glatstein, M.D.
Radiation Oncology Branch, National Cancer Institute, Bethesda, Maryland 20014

INTRODUCTION

The diversity of the non-Hodgkin's lymphomas challenges all physicians who participate in the diagnosis, staging, and therapy of this group of disorders. In the following discussion, emphasis is given to our current knowledge of histopathologic classification, staging of disease extent, and the roles of radiation therapy, chemotherapy, and combined modality treatments. Since there are few, if any, universally accepted approaches to these diseases, the data presented in this review reflect that controversy.

PATHOLOGY

The non-Hodgkin's lymphomas comprise a broad spectrum of lymphoid malignancies with different natural histories, histopathologies, and responses to therapy. The terms reticulum cell sarcoma, lymphosarcoma, and giant follicular lymphoma have proved to be of little use to the clinician, because they are inadequate for correlating natural history of disease or response to treatment. On the other hand, the Rappaport classification (1) is an effective clinical tool, identifying important distinctions in prognosis between the generally indolent nodular or follicular lymphomas and the more aggressive diffuse lymphomas. In addition to architectural patterns, the Rappaport scheme describes the malignant cell population as either lymphocytic, histiocytic, or mixed lymphocytic-histiocytic (Table 1). And here too, there is prognostic value, for within both the nodular and diffuse lymphomas, a lymphocytic cytology is more favorable than a histiocytic one.

0066-4219/78/0401-0081$01.00

Unfortunately, even this relatively simple classification requires expert hematopathologic interpretation. A recent analysis from the Southwest Oncology Group (2) showed that there was a marked discrepancy between the interpretation of biopsies by institutional pathologists as compared to experts of the Lymphoma Panel. Most dissatisfaction with the Rappaport scheme, however, resulted from the recognition that many lymphomas, termed histiocytic on morphologic grounds alone by Rappaport, have recently been shown by ultrastructural and immunologic techniques to be neoplasms of large or transformed lymphocytes rather than of histiocytic origin (3–6). Hence, new classification systems have been proposed based on functional studies (7–10). Although all agree that any classification system should be scientifically accurate as well as clinically relevant, none of the newer classification systems have yet been established as practically superior to the Rappaport scheme. Moreover, it is not clear that any one of the newer systems can be reproduced with any more accuracy among pathologists of varying skills than the Rappaport classification. Detailed information concerning the clinicopathologic correlations of the newer classifications is largely fragmentary.

Another serious problem in classifying the non-Hodgkin's lymphomas is that as many as 20% of patients may have multiple histologies, and occasionally, there are even "composite" lymphomas in which two kinds of lymphoma occur in different areas within the same lymph node (11). In addition, an involved lymph node may frequently reveal both a partly nodular and diffuse architecture, while containing a single cell type. At present there is conflicting information concerning the importance of the degree of nodularity for prognosis in such patients. Patchevsky et al (12) reported that the greater the degree of nodularity detected microscopically, the better the prognosis, whereas Warnke et al (13) demonstrated that patients with poorly differentiated lymphocytic lymphoma of both nodular and diffuse architecture in the same lymph node have a prognosis that is not significantly different from those patients with a nodular architecture alone. Likewise, patients with both nodular and diffuse mixed lymphoma in the same lymph node have an expected survival and disease-free survival rate similar to those patients whose biopsy shows only a nodular mixed pattern. In patients with NHL those with both nodular and diffuse architecture appear to have a less favorable prognosis than those with a nodal architecture which appears entirely nodular. However, the number of patients with nodular his-

Table 1 The Rappaport classification of non-Hodgkin's lymphomas

	Architecture	
Cytology	Nodular	Diffuse
Lymphocytic, well differentiated (LWD)	NLWD	DLWD
Lymphocytic, poorly differentiated (LPD)	NLPD	DLPD
Mixed lymphocytic and histiocytic (ML)	NML	DML
Histiocytic (HL)	NHL	DHL

tiocytic lymphoma is quite small and, accordingly, the data are more difficult to interpret.

STAGING

The optimal system for staging the non-Hodgkin's lymphomas remains to be defined. At present, the Ann Arbor staging system designed for Hodgkin's disease is also employed in the non-Hodgkin's lymphomas (14). However, its clinical utility is less evident for non-Hodgkin's lymphomas, probably because the survival of most patients with nodular lymphoma is prolonged even when bone marrow disease is evident at presentation, and in patients with diffuse lymphomas, the survival tends to be comparatively short, even with nodal disease alone.

One of the major differences at presentation between patients with Hodgkin's disease and those with other malignant lymphomas is the marked tendency for extranodal involvement in the non-Hodgkin's lymphomas (11). As a result, much more attention must be paid to such areas as bone, liver, GI tract, Waldeyer's ring, and bone marrow. The latter is undoubtedly the most important single organ to evaluate thoroughly prior to the initiation of treatment, and bone marrow biopsy, with or without aspiration, is an essential part of the staging procedure. Often multiple biopsies are desirable, especially in patients with lymphocytic lymphoma (15). The use of lymphography and other radiologic procedures and scans is important in the evaluation of any patient who is a likely candidate for aggressive and potentially curative treatment. With a thorough staging evaluation, almost 70% of patients with non-Hodgkin's lymphomas will present with Stage III or IV disease (16, 17).

More aggressive staging techniques, including peritoneoscopy (17) and staging laparotomy, have been advocated by several centers. To date, the major observation from laparotomy staging is the frequent (50%) recognition of mesenteric lymph node involvement, usually unsuspected on clinical grounds (16). Such documented involvement has necessitated major technical changes in radiotherapy portals to the abdomen (18). However, it is now clear that staging laparotomy remains an investigational tool and not a routine procedure. The usefulness of surgical staging will depend to a large degree on the plan and effectiveness of treatment.

RADIATION THERAPY

Relatively few data are available on the effectiveness of varying radiation doses to sterilize malignant lymphomas classified according to Rappaport. In a comprehensive study, Fuks & Kaplan (19) showed that progressive doses in excess of 4000 rad achieved very high local control rates in patients with nodular lymphoma. In a different analysis, Cox et al (20) questioned the need to go above 3000 rad; however, their data have been challenged (21) on the basis of inadequate restaging when patients were thought to be in clinical remission. For patients with diffuse lymphoma, Fuks & Kaplan (19) noted local recurrences in approximately 15% of fields at risk, even with doses in excess of 4000 rad. Moreover, although most of the diffuse subtypes showed a dose response, DHL did not. The explanation for this is somewhat unclear, but probably reflects the fact that more than one type of cell can appear to be histiocytic with light microscopy, and that many large tumors that did not respond to modest doses of radiation were not excluded from their analysis.

In addition to dose, the importance of volume that is irradiated has generally been neglected in the literature. For clinical Stage I and II non-Hodgkin's lymphomas, the results of involved field or regional lymph node radiation have been reported by only a few authors who used the Rappaport classification (22, 23). In these series, it is apparent that Stage I disease can be successfully controlled with local or regional radiation therapy in slightly more than half the patients, assuming that there has been adequate staging, which includes lymphography and bone marrow biopsy. Stage II patients have fared less well, with disease-free survival rates of 25–45%.

Data concerning the sites of relapse for Stage I and II disease treated with radiotherapy alone are scant. Fuks et al (24) demonstrated a significant percentage of patients who failed initially in extranodal sites; however, even more frequently in Stage I and II patients, nonirradiated lymph nodes proved to be the initial site of relapse after local or extended field radiotherapy.

Total lymphoid radiotherapy, i.e. radiation therapy to encompass major lymph node groups on both sides of the diaphragm, has seldom been used in the treatment of Stage I or II non-Hodgkin's lymphomas. Glatstein et al (25) demonstrated a 65–70% relapse-free survival rate at five years after total lymphoid irradiation in patients with unfavorable histologies (NHL, DLPD, DML, DHL) of Stage I and II extent. Excellent results in Stage I and II DHL, pathologically staged and treated with radiotherapy to the mantle and to prophylactic fields on the opposite side of the diaphragm, were reported by Bitran et al (26) as well. These authors emphasized the importance of the number of involved sites of Stage II DHL: There were no failures among patients who had fewer than four sites of involvement, while patients who had four or more involved sites had a high likelihood of relapse when treated with radiotherapy alone.

For Stage I and II favorable non-Hodgkin's lymphomas (NLPD, NML, DLWD), there is little evidence to support the use of total lymphoid radiotherapy in patients who have been pathologically staged. It is important to stress, however, that the incidence of unsuspected subdiaphragmatic disease in mesenteric lymph nodes is approximately 40% in those with favorable histologies and a negative lymphogram (16). Thus, if radiation therapy is to be administered with curative intent and surgical staging has not been performed, a strong case can be made for encompassing the entire abdominal contents.

Radiotherapy alone for Stage III lymphomas of unfavorable histologies has yielded generally poor results. On the other hand, total lymphoid irradiation alone in Stage III nodular lymphomas resulted in remission durations of up to 10 years in one third of patients, and survival data that are comparable to other modalities (27).

In recent years, there has been renewed interest in whole body irradiation (28–30). Despite its obscure mechanism of action, this method of irradiation has achieved survival results that are similar to other kinds of treatment without much toxicity other than delayed hematologic depression. This latter side effect may compromise later chemotherapy, although it has been claimed not to do so (31). Data on freedom from relapse suggests that whole body irradiation is an effective tool for palliation, rather than cure.

CHEMOTHERAPY

Chemotherapy plays a major role in the treatment of the non-Hodgkin's lymphomas, since the disease is frequently of extranodal extent in the favorable lymphomas (NLPD, NML, DLWD) and, even when localized to lymph nodes (Stage II and III) in the unfavorable histologies (NHL, DLPD, DML, DHL), results with radiation therapy alone are generally unsatisfactory. Second, many drugs are highly effective as single agents and as part of combination chemotherapy programs in achieving complete or partial regressions in these diverse diseases. In a retrospective study of patients treated with single-agent chemotherapy, Jones et al (32) pointed out that both a nodular architecture and a lymphocytic cytology were correlated with a significantly greater likelihood of response to single-agent therapy and prolonged survival as compared to the diffuse or histiocytic lymphomas. Clinical complete responses were seen in 48% of patients with NLPD in contrast to only 5% of those with DHL, and responses in the other cytologic and architectural patterns were distributed between these two extremes (NML, 31%; NHL, 28%; DLPD, 22%; DML, 13%). Survival was likewise superior for patients with NLPD where the median survival was 60+ months compared to less that 7.5 months for DHL.

Others, using combination chemotherapy, observed similar response and survival correlations with histopathology (33–38). Schein et al (33) treated 80 patients with CVP, MOPP, or C-MOPP[1] and found significantly more pathologically documented complete responses in the nodular lymphomas (NLPD, 60%; NML, 71%) as compared to the diffuse lymphomas (DLWD, 50%; DLPD, 22%; DML, none; DHL, 34.6%). And again, overall survival duration in each histologic subgroup reflected the favorable determinants of a nodular architecture and a lymphocytic cytology (median survival for NLPD, 40+ mo; DHL, 6 mo).

Whether combination chemotherapy is superior to single-agent treatment in advanced, favorable lymphomas is not yet established. Pathologically documented complete responses are reported in 60–90% of patients treated with combination chemotherapy programs (33–38). Similarly, in a prospective randomized trial comparing single–alkylating-agent therapy to CVP or to split-course CVP and total lymphoid irradiation, 65% of single-agent patients obtained pathologically documented complete remissions and there were no significant differences in complete response rates between the three groups (39). The survival in the three-armed study was excellent, with more than 80% of all patients alive at five years. Kennedy et al (40) reported similar results comparing sequential cyclophosphamide, vincristine, and prednisone to combination CVP.

---

[1] CVP: cyclophosphamide, vincristine, prednisone; MOPP: nitrogen mustard, vincristine, procarbazine, prednisone; C-MOPP: cyclophosphamide, vincristine, procarbazine, prednisone.

Many groups observed that in general the complete remissions obtained with combination chemotherapy in advanced, favorable lymphomas are usually not durable despite prolonged survival time (33, 35–38). Schein et al (41) described a pattern of continuous, late recurrence in patients with NLPD and DLWD, with a median disease-free survival time of approximately two years; yet their overall survival rate was more than 80% at 40 mo. Schein observed (41), and recently it has been emphasized by Anderson et al (42), that complete responders with NML do not appear to have the same pattern of continuous relapse as those with NLPD. In contrast to NLPD, where the median remission duration was 15.8 mo, the median was in excess of seven years for NML. However, in spite of these apparent differences in remission duration, the median survival time of both groups was greater than six years. As a result, conclusions regarding the impact of remission duration on survival time in these favorable lymphomas cannot yet be drawn.

Because of the indolent nature of these advanced diseases, some have asked whether initial treatment is necessary in asymptomatic patients. Portlock & Rosenberg (39) reported 35 previously untreated, relatively asymptomatic patients with advanced, favorable lymphomas who were closely followed without initial therapy. Half of these patients required some form of treatment (usually conservative) within three years, while 16 were untreated for periods of up to ten years. This selected group of patients had an overall survival rate of 70% at 10 years.

Future clinical trials in the advanced, favorable lymphomas will need to address the following questions: (*a*) Is any chemotherapy regimen superior to any other in terms of response rate, remission duration, and survival? (*b*) How can remission duration be improved? (*c*) What is the impact of remission duration on survival in these indolent diseases? (*d*) Can the advanced, favorable lymphomas be cured?

In advanced, diffuse histiocytic lymphoma, the impact of combination chemotherapy has been dramatically demonstrated. Unlike the single–alkylating-agent complete response rate of only 5% in DHL, many groups obtained complete remission rates of 40–93% with combination chemotherapy (33, 36–39, 43–47). More importantly, in contrast to the rapidly progressive and invariably fatal course of DHL in patients treated with single alkylating agents, prolonged disease-free survival and probable cure has been reported in as many as 40% of patients using combination chemotherapy (43, 47).

Unlike NLPD where one observes a pattern of continuous, late relapse from complete remission, relapse of DHL following pathologically documented complete response is uncommon, and usually occurs within the first two years after treatment (41, 43). The remarkable durability of these complete remissions is evidenced by reports of disease-free survivors alive for more than five years after chemotherapy alone (43, 44). Thus, the relative efficacy of chemotherapy combinations for advanced DHL must be judged not only by complete response rate, but more importantly by remission duration. To date, there are two regimens used in previously untreated patients that appear comparable in terms of these parameters: MOPP or C-MOPP (43) and BACOP[2] (47). Both reported complete response rates of more than 40% with only one relapse among 23 complete responders, who had been off all therapy for as long as eight years. Others have reported similar or better complete response rates, but with less impressive remission duration (36–39, 44, 46). The reason for these apparent differences in remission duration can probably be attributed to the definition of complete response. When careful restaging is done one month after the completion of all chemotherapy, as many as 25% of patients with a clinical complete response will have residual microscopic disease and subsequently develop clinically apparent relapse (47). Furthermore, it is well recognized that a partial response in advanced DHL conveys little survival benefit. The median lifespan is usually less than eight months regardless of the degree of partial response. Therefore, what factors may contribute to an improved, complete remission rate and survival? Previously untreated patients generally fare better than those with prior radiation therapy, and, in particular, prior chemotherapy. Disease extent also influences survival rate, as reported by Fisher et al (48). With Stage III DHL, 66% of patients treated with C-MOPP or BACOP were alive at three years, whereas with extranodal involvement the median survival was only 9.5 mo. Also, large tumors were associated with a poor response to drugs, whereas differences in age, sex, and constitutional symptoms were not.

Meningeal disease as the only site of DHL relapse in clinical complete responders has recently been reported (49), but there is not enough information at present to answer the question of whether CNS prophylaxis is warranted on a routine basis.

Chemotherapy results for the other unfavorable histologies — DLPD, DML, and NHL — are fragmentary. In general, aggressive combination chemotherapy has been disappointing in DML; few patients have obtained complete remissions (33, 35, 47). Treatment results in NHL are not available. DLPD represents at least two diseases (DLPD and lymphoblastic lymphoma) (50), and consequently results of therapy are difficult to assess. Complete responses have been reported in 22–80% of patients and some patients appear to have a prolonged disease-free survival time, although the numbers are small (33, 35, 38).

Future chemotherapy questions in the unfavorable lymphomas include the following: (*a*) How can the pathologically documented complete remission rate in advanced DHL be maximized? (*b*) Is there a role for chemotherapy alone in Stage II DHL? (*c*) What is the relative risk of CNS relapse, and is there a need for prophylaxis? (*d*) How can results in the other unfavorable histologies (DLPD, DML, and NHL) be improved?

---

[2] BACOP: bleomycin, adriamycin, cyclophosphamide, vincristine, prednisone.

COMBINED MODALITY TREATMENT

Since the most likely sites of failure after chemotherapy appear to be previous sites of known involvement (41), and the most likely sites of failure after radiotherapy appear to be nonirradiated lymph nodes and extranodal sites (24), it is surprising that few studies have been reported that utilize both radiotherapy and chemotherapy in the initial treatment of patients with non-Hodgkin's lymphomas.

Preliminary results have been published by Monfardini et al (51, 52) treating Stage I and II non-Hodgkin's lymphomas of all histologies with involved-field radiotherapy, with or without six cycles of adjuvant chemotherapy (CVP). Patients were not randomized to CVP until they had successfully completed the entire course of radiotherapy; significant improvement is reported in relapse-free survival and survival for patients who received the adjuvant chemotherapy.

A similar study (25) in patients with Stage I and II unfavorable lymphomas showed no advantage of receiving adjuvant chemotherapy as compared to treatment with radiotherapy alone. In this study, however, radiation therapy was administered to all lymph node groups. For Stage III and IV unfavorable lymphomas, combined modality treatment appeared beneficial when compared to chemotherapy alone; however, the results were not striking. Preliminary results combining chemotherapy with whole body irradiation have been reported and appear to be encouraging (53).

The benefits of the combined modality approach are just beginning to be tested. A major problem remaining to be solved is how to combine chemotherapy and radiation therapy maximally without compromising the efficacy of either treatment.

CONCLUSION

With current treatment methods, the majority of patients with non-Hodgkin's lymphomas can obtain a pathologically documented complete remission, and in some histologic subtypes, in particular diffuse histiocytic lymphoma, the potential for durable complete remissions and actual cure is well established. Future goals of treatment must look not only to improving the frequency and duration of complete remissions, but also to demonstrating the impact of relapse-free survival on the survival of patients with all histologic subtypes of disease.

ACKNOWLEDGMENT

The authors wish to thank Ms. Nancy MacLachlan for her secretarial assistance.

Literature Cited

1. Rappaport, H., Winter, W. J., Hicks, E. B. 1956. Follicular lymphoma — A re-evaluation of its position in the scheme of malignant lymphoma, based on a survey of 253 cases. *Cancer* 9:792–821
2. Jones. S. E., Butler, J. J., Byrne, G. E., Coltmann, C. A., Moon, T. E. 1977. Histopathologic review of lymphoma cases from the Southwest Oncology Group. *Cancer* 39:1071–76
3. Levine, G. D., Dorfman, R. F. 1975. Nodular lymphoma: An ultrastructural study of its relationship to germinal centers and a correlation of light and electron microscopic findings. *Cancer* 35:148–64
4. Glick, A. D., Leech, J. H., Waldron, J. A., Flexner, J. M., Horn, R. G., Collins, R. D. 1975. Malignant lymphomas of follicular center cell origin in man. II. Ultrastructural and cytochemical studies. *J. Natl. Cancer Inst.* 54:23–36
5. Jaffe, E. S., Shevach, E. M., Frank, M. M., Berard, C. W., Green, I. 1974. Nodular lymphoma — Evidence for origin from follicular B lymphocytes. *N. Engl. J. Med.* 290:813–19
6. Leech, J. H., Glick, A. D., Waldron, J. A., Flexner, J. M., Horn, R. G., Collins, R. D. 1975. Malignant lymphomas of follicular center cell origin in man. I. Immunologic studies. *J. Natl. Cancer Inst.* 54:11–21
7. Lukes, R. J., Collins, R. D. 1974. Immunologic characterization of human malignant lymphomas. *Cancer* 34:1488–1503
8. Dorfman, R. F. 1974. Classification of non-Hodgkin's lymphomas. *Lancet* 1:1295–96
9. Bennett, M. H., Farrer-Brown, G., Henry, K., Jelliffe, A. M. 1974. Classification of non-Hodgkin's lymphomas. *Lancet* 2:405–6
10. Gerard-Marchant, R., Hamlin, I., Lennert, K., Rilke, F., Stansfeld, A. G., van Unnik, J. A. M. 1974. Classification of non-Hodgkin's lymphomas. *Lancet* 2:406–8
11. Goffinet, D. R., Castellino, R. A., Kim, H., Dorfman, R. F., Fuks, Z., Rosenberg, S. A., Nelsen, T., Kaplan, H. S. 1973. Staging laparotomies in unselected previously untreated patients with non-Hodgkin's lymphomas. *Cancer* 32:672–81
12. Patchefsky, A. S., Brodovsky, H. S., Menduke, H., Southard, M., Brooks, J., Nicklas, D., Hoch, W. S. 1974. Non-Hodgkin's lymphomas: A clinicopathologic study of 293 cases. *Cancer* 34:1173–86
13. Warnke, R. A., Kim, H., Fuks, Z., Dorfman, R. F. 1977. The coexistence of nodular and diffuse patterns in nodular non-Hodgkin's lymphomas: Significance and clinicopathologic correlation. *Cancer.* In press
14. Carbone, P. P., Kaplan, H. S., Musshoff, K., Smithers, D. W., Tubiana, M. 1971. Report of the committee on Hodgkin's Disease Staging Classification. *Cancer Res.* 31:1860–61
15. Rosenberg, S. A. 1975. Bone marrow involvement in the non-Hodgkin's lymphomata. *Br. J. Cancer* 31 (Suppl. II):261–64
16. Goffinet, D. R., Warnke, R., Dunnick, N. R., Castellino, R., Glatstein, E., Nelsen, T. S., Dorfman, R. F., Rosenberg, S. A., Kaplan, H. S. 1977. Clinical and surgical (laparotomy) evaluation of patients with non-Hodgkin's lymphomas. *Cancer Treat. Rep.* In press
17. Chabner, B. A., Johnson, R. E., Young, R. C., Canellos, G. P., Hubbard, S. P., Johnson, S. K., De Vita, V. T. Jr. 1976. Sequential nonsurgical and surgical staging of non-Hodgkin's lymphoma. *Ann. Intern. Med.* 85:149–54
18. Goffinet, D. R., Glatstein, E., Fuks, Z., Kaplan, H. S. 1976. Abdominal irradiation in non-Hodgkin's lymphomas. *Cancer* 37:2797–2806
19. Fuks, Z., Kaplan, H. S. 1973. Recurrence rates following radiation therapy of nodular and diffuse malignant lymphomas. *Radiology* 108:675–84
20. Cox, J. D., Koehl, R. H., Turner, W. M., King, F. M. 1974. Irradiation in the local control of malignant lymphoreticular tumors (non-Hodgkin's malignant lymphoma). *Radiology* 112:179–85
21. Glatstein, E. 1976. Alice in Lymphoma-land. *Int. J. Radiat. Oncol. Biol. Phys.* 1:561–63
22. Jones, S. E., Fuks, Z., Kaplan,

H. S., Rosenberg, S. A. 1973. Non-Hodgkin's lymphomas. V. Results of radiotherapy. *Cancer* 32:682–91

23. Hellman, S., Chaffey, J. T., Rosenthal, D. S., Moloney, W. C., Canellos, G. P., Skarin, A. T. 1977. The place of radiation therapy in the treatment of non-Hodgkin's lymphomas. *Cancer* 39:843–51
24. Fuks, Z., Glatstein, E., Kaplan, H. S. 1975. Patterns of presentation and relapse in the non-Hodgkin's lymphomata. *Br. J. Cancer* 31 (Suppl. II):286–97
25. Glatstein, E., Donaldson, S. S., Rosenberg, S. A., Kaplan, H. S. 1977. The potential for combined modality therapy in malignant lymphomas. *Cancer Treat. Rep.* In press
26. Bitran, J. D., Kinzie, J., Sweet, D. L., Variakojis, D., Griem, M. L., Golomb, H. M., Miller, J. B., Oetzel, N., Ultmann, J. E. 1977. Survival of patients with localized histiocytic lymphoma. *Cancer* 39:342–46
27. Glatstein, E., Fuks, Z., Goffinet, D. R., Kaplan, H. S. 1976. Non-Hodgkin's lymphomas of Stage III extent: Is total lymphoid irradiation appropriate treatment? *Cancer* 37:2806–12
28. Johnson, R. E., O'Conor, G. T., Levin, D. 1970. Primary management of advanced lymphosarcoma with radiotherapy. *Cancer* 25:787–91
29. Johnson, R. E. 1975. Management of generalized malignant lymphomata with "systemic" radiotherapy. *Br. J. Cancer* 31 (Suppl. II):450–55
30. Chaffey, J. T., Rosenthal, D. S., Moloney, W. C., Hellman, S. 1976. Total body irradiation as treatment for lymphosarcoma. *Int. J. Radiat. Oncol. Biol. Phys.* 1:399–405
31. Brereton, H. D., Fairey, R. N., Johnson, R. E. 1977. Successful chemotherapy of advanced non-Hodgkin's lymphoma after relapse following total body irradiation. *Int. J. Radiat. Oncol. Biol. Phys.* 2:249–55
32. Jones, S. E., Rosenberg, S. A., Kaplan, H. S., Kadin, M. E., Dorfman, R. F. 1972. Non-Hodgkin's lymphomas. II. Single agent chemotherapy. *Cancer* 30:31–38
33. Schein, P. S., Chabner, B. A., Canellos, G. P., Young, R. C., Berard, C., De Vita, V. T. Jr. 1974. Potential for prolonged disease-free survival following combination chemotherapy of non-Hodgkin's lymphoma. *Blood* 43:181–89
34. Skarin, A. T., Pinkers, G. S., Myerowitz, R. L., Bishop, Y. M., Moloney, W. C. 1974. Combination chemotherapy of advanced lymphocytic lymphoma: Importance of histologic classification in evaluating response. *Cancer* 34:1023–29
35. Portlock, C. S., Rosenberg, S. A. 1976. Combination chemotherapy with cyclophosphamide, vincristine, and prednisone in advanced non-Hodgkin's lymphomas. *Cancer* 37:1275–82
36. McKelvey, E. M., Gottlieb, J. A., Wilson, H. E., Haut, A., Talley, R. W., Stephens, R., Lane, M., Gamble, J. F., Jones, S. E., Grozea, P. N., Gutterman, J., Coltman, C. Jr., Moon, T. E. 1976. Hydroxyldaunomycin (adriamycin) combination chemotherapy in malignant lymphoma. *Cancer* 38:1484–93
37. Rodriguez, V., Cabanillas, F., Burgess, M. A., McKelvey, E. M., Valdivieso, M., Bodey, G. P., Freireich, E. J. 1977. Combination chemotherapy ("CHOP-Bleo") in advanced (non-Hodgkin) malignant lymphoma. *Blood* 49:325–33
38. Skarin, A. T., Rosenthal, D. S., Moloney, W. C., Frei, E. III. 1977. Combination chemotherapy of advanced non-Hodgkin lymphoma with bleomycin, adriamycin, cyclophosphamide, vincristine, and prednisone (BACOP). *Blood* 49:759–68
39. Portlock, C. S., Rosenberg, S. A. 1977. Chemotherapy of the non-Hodgkin's lymphomas: The Stanford experience. *Cancer Treat. Rep.* In press
40. Kennedy, B. J., Hill, J., Bloomfield, C., Kiang, D., Fortuny, I., Theologides, A. 1975. Combination (COP) versus successive single agent chemotherapy (C-O-P) in lymphocytic lymphoma. *Proc. Am. Soc. Clin. Oncol.* 16:142 (Abstr.)
41. Schein, P. S., Chabner, B. A., Canellos, G. P., Young, R. C., De Vita, V. T. Jr. 1975. Non-Hodgkin's lymphoma: Patterns of relapse from complete remission after combination chemotherapy. *Cancer* 35:354–57
42. Anderson, T., De Vita, V., Bender, R., Mann, R., Chabner, B., Young, R. 1977. Nodular lymphomas (NL):

An important correlation of survival with complete remission induction. *Proc. Am. Soc. Clin. Oncol.* 18:326 (Abstr.)

43. De Vita, V. T. Jr., Canellos, G. P., Chabner, B., Schein, P., Hubbard, S. P., Young, R. C. 1975. Advanced diffuse histiocytic lymphoma, a potentially curable disease: Results with combination chemotherapy. *Lancet* 1:248–50
44. Cadman, E., Berd, D., Farber, L., Bertino, J. R. 1977. Combination therapy of diffuse histiocytic lymphoma that includes antimetabolites. *Cancer Treat. Rep.* In press
45. Lauria, F., Baccarani, M., Fiacchini, M., Gobbi, M., Tura, S. 1975. Methotrexate, cyclophosphamide and vincristine (MEV regimen) for non-Hodgkin's lymphoma. *Eur. J. Cancer* 11:343–49
46. Durant, J. R., Loeb, V. Jr., Dorfman, R., Chan, Y. K. 1975. 1,3-Bis-(2-chloroethyl)-1-nitrosourea (BCNU), cyclophosphamide, vincristine and prednisone (BCOP): A new therapeutic regimen for diffuse histiocytic lymphoma. *Cancer* 36:1936–44
47. Schein, P. S., De Vita, V. T. Jr., Hubbard, S., Chabner, B. A., Canellos, G. P., Berard, C., Young, R. C. 1976. Bleomycin, adriamycin, cyclophosphamide, vincristine, and prednisone (BACOP) combination chemotherapy in the treatment of advanced diffuse histiocytic lymphoma. *Ann. Intern. Med.* 85:417–22
48. Fisher, R. I., Johnson, B. L., Simon, R., Schein, P. S., De Vita, V. T., Young, R. C. 1976. Factors influencing survival of diffuse histiocytic lymphoma (DHL) following combination chemotherapy. *Proc. Am. Soc. Clin. Oncol.* 17:257 (Abstr.)
49. Bunn, P. A., Schein, P. S., Banks, P. M., De Vita, V. T. Jr. 1976. Central nervous system complications in patients with diffuse histiocytic and undifferentiated lymphoma: Leukemia revisited. *Blood* 47:3–10
50. Nathwani, B. N., Kim, H., Rappaport, H. 1976. Malignant lymphoma, lymphoblastic. *Cancer* 38:964–83
51. Monfardini, S., Lattuada, A., Milani, F., Banfi, A., Bonadonna, G. 1977. Controlled study with radiotherapy (RT) vs RT plus CVP in stage I-II non-Hodgkin's lymphomas. *Proc. Am. Soc. Clin. Oncol.* 18:283 (Abstr.)
52. Lattuada, A., Bonadonna, G., Milani, F., Banfi, A., Valagussa, P., De Lena, M., Monfardini, S. 1977. Adjuvant chemotherapy with CVP after radiotherapy (RT) in stage I-II non-Hodgkin's lymphomas. In *Adjuvant Therapy of Cancer,* ed. S. E. Salmon, S. E. Jones, pp. 537–44. Amsterdam: Elsevier/North-Holland. 646 pp.
53. Harrison, D. T., Neiman, P. E. 1977. Combined modality therapy in advanced non-Hodgkin's lymphoma of aggressive histologic type. *Proc. Am. Soc. Clin. Oncol.* 18:323 (Abstr.)

Ann. Rev. Med. 1978. 29:93–98
Copyright © 1978 by Annual Reviews Inc. All rights reserved

DIALYSIS ENCEPHALOPATHY SYNDROME

◆7269

Allen C. Alfrey, M.D.
Renal Division, Denver Veterans Administration Hospital and University of Colorado Medical Center, Denver, Colorado 80220

In 1972, a distinct clinical syndrome of a progressive fatal encephalopathy that occurred in five chronic hemodialysis patients was reported (1). Although this syndrome was initially thought to be restricted to the Denver area, it has subsequently become apparent that its distribution is widespread and that it is a significant cause of death in the dialysis population (2–4). A recent review of deaths in 6 dialysis units carried out by Schreeder (5) showed that, of 574 deaths, 61 or 10.6% were a result of this disease. In the various units studied, the percentage of deaths from this syndrome varied from 4 to 35%. Of the 114 deaths that occurred in patients who had been maintained on dialysis for periods in excess of 3 years, 20 or 17.5% of the patients died of this disease. Although the units studied were not selected randomly, the above findings suggest that this syndrome may occur with considerably greater frequency than previously thought.

CLINICAL FEATURES

The first symptom is usually a mild cortical speech difficulty characterized by a stammering, hesitancy of speech, and, at times, speech arrest. The speech disorder is intensified during and immediately following dialysis and may initially be seen only during these periods. In association with the speech disorder, there are usually subtle mental and psychiatric changes. In conjunction with these alterations is a consistent electroencephalographic (EEG) abnormality. The EEG changes may precede any clinical findings by six to eight months. The characteristic EEG changes are mild slowing of the dominant rhythm with increased low voltage theta and bursts of anteriorly predominant high voltage delta (Figures 1 and 2). At times, the paroxysmal activity may take the form of a 1.3–3 c/sec spike and wave. These changes are usually located in the front-ocentral regions. The EEG changes differ from those of uremia and a number of other metabolic encephalopathies by the relatively normal background frequency. Once EEG changes are manifested, they tend to persist throughout the

0066-4219/78/0401-0093$01.00

course of the illness. In my experience, the classical EEG changes are uniformly present in patients with this syndrome, and, without such changes, the diagnosis of the dialysis-associated encephalopathy syndrome cannot be made.

As the disease progresses the speech disturbances become increasingly severe, manifested as a combination of dysarthria, dyspraxia, and dysphasia. There is also a progressive global dementia and, at times, prominent parietal lobe signs such as directional disorientation and construction apraxia. Personality changes become marked with paranoid thinking, psychoses, and even delirium. In addition the patients frequently develop seizures, asterixis, twitching, motor apraxia, and myoclonic jerks. The usual course of the disease is 6–12 months during which time the patients lose their ability to communicate and ambulate and eventually are unable to perform any purposeful movements. The clinical findings in 12 patients are listed in Table 1. Major causes of death include sepsis, suicide, and removal from the dialysis program because of neurological and psychiatric disabilities. Because of the high incidence of psychiatric disturbances and suicide in these patients, this diagnosis should be entertained in any dialysis patient who has a marked change in personality or attempts suicide.

A number of factors tend to unmask this syndrome. As stated above, symptoms may be initially evident in the early postdialysis period and, once the disease is fully manifested, it is commonly aggravated during and immediately following dialysis. The syndrome can also occur in association with surgical trauma, infection, and during hypercalcemic episodes (6). With correction of these conditions, the clinical features may totally clear, with the exception of the EEG changes, only to reccur in 6–12 months and persist thereafter until the patient dies.

With the exception of the EEG, other laboratory tests are of little help in making the diagnosis and are largely carried out to exclude other neurological diseases. Lumbar puncture and cerebrospinal fluid are normal, as are skull films and brain scans. Computerized axial tomography in a number of patients with this syndrome has shown mild to moderate hydrocephalus *ex vacuo* which supports the presence of cortical atrophy.

Table 1 Clinical findings in 12 patients

Symptoms/signs	Patients (no.)	Patients (%)
Dementia	12	100
Speech Difficulties	12	100
Markedly abnormal EEG	12	100
Seizures	9	75
Twitching or myoclonic jerks	8	75
Asterixis	5	42
Focal neurological deficits	5	42
Peripheral neuropathy	5	42
Apraxia	4	33

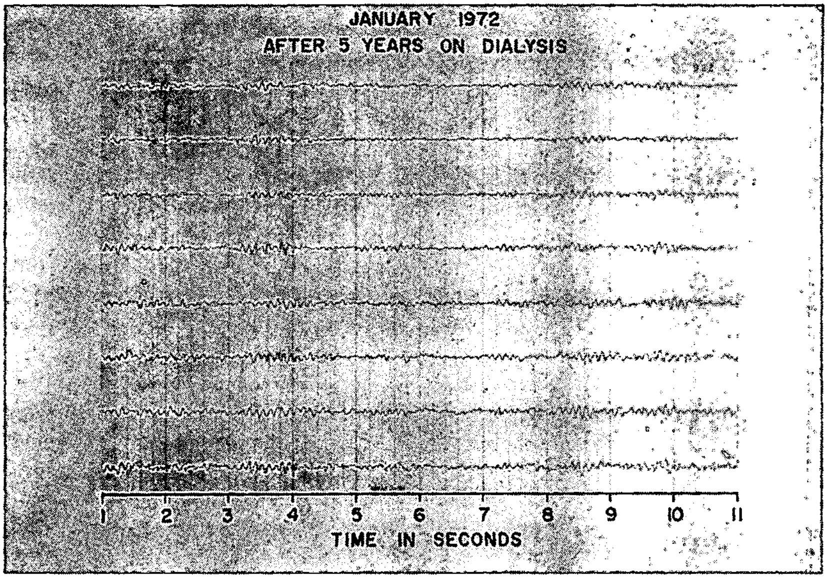


Figure 1 EEG tracing obtained in a patient who had received intermittent hemodialysis for five years.

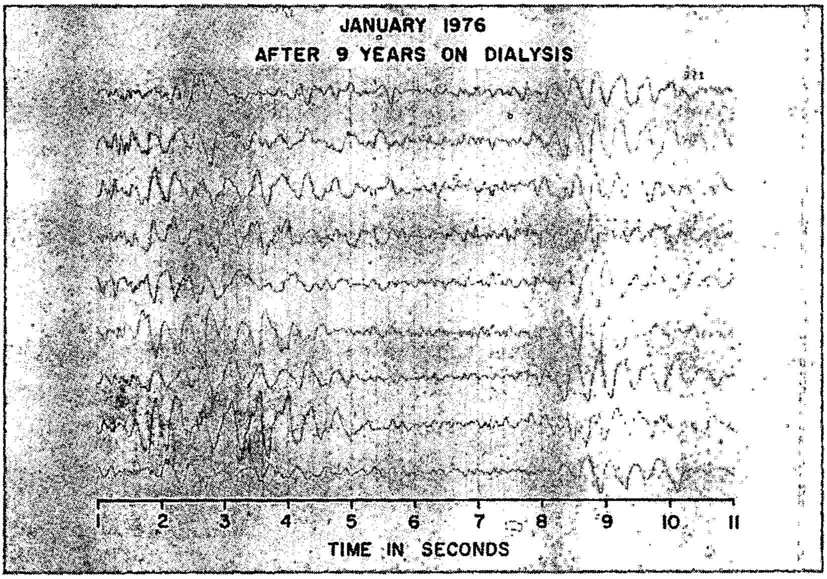


Figure 2 EEG tracing obtained on the same patient four years later after developing the dialysis-associated encephalopathy syndrome.

ETIOLOGIES

A number of different etiological possibilities have been suggested; however, the three major considerations are normal pressure hydrocephalus, slow virus infection, and toxin-induced encephalopathy.

On the basis of abnormal cysternograms as determined with radioiodinated serum albumin in several of these patients, Mahurkar and co-workers (7) suggested that the disease was a result of deranged cerebrospinal fluid dynamics similar to that found in normal pressure hydrocephalus. However, other findings have not supported this suggestion; brains examined at autopsy have not shown hydrocephalus (8) and computerized axial tomography of the brain has revealed cortical atrophy rather than internal hydrocephalus.

Because of the clinical similarities between the dialysis encephalopathy syndrome and Creutzfeldt-Jacob syndrome, it has been suggested that the former condition may be a slow virus infection (9). Although histological changes including shrinkage of neurons, increased paired astrocytes, microglial changes, and spongy changes in the second and third cortical layers have been described in patients who died of this syndrome, these changes are no different from those found in the brains of dialysis patients dying of other causes (8). In addition, the spongy changes are inconsistent and not as severe as those found in the Creutzfeldt-Jacob disease (10).

The clinical features, EEG changes, lack of distinctive histological changes, and apparent geographical variation in incidence of this entity all suggest the possibility of a neurotoxin being the etiological agent. Although earlier studies suggested that tin might play a role in the etiology of this condition (1), additional studies have shown that tin levels are not consistently increased in brains of patients dying from this condition and, in fact, tin may be present in concentrations equal to or higher in dialysis patients who die of other causes. Only one element, aluminum, has been found to be consistently elevated in the brains of patients dying of the dialysis-associated encephalopathy syndrome. This finding was first reported from our laboratory in 1976 (11). Subsequently, two additional reports have appeared in the literature documenting increased brain aluminum levels in patients with this condition (10, 12). In our initial report, we studied 6 brains from patients who died of this condition; we have now increased that number to 25. There continues to be little overlap between dialysis patients who die of this syndrome and those who die of other causes. The lowest brain aluminum level found to date in any patient with this syndrome was 12 mg/kg dry weight gray matter (normal is 2 mg/kg).

The environmental source responsible for the alteration in brain aluminum levels in patients dying of the encephalopathy syndrome has not been clearly defined. The two most likely sources are some absorption of aluminum from the orally administered phosphate-binding agents and aluminum contamination in the dialysate. A recent study has documented that, in normal individuals given aluminum-containing antacids, urinary aluminum excretion increases by a mean value of 500 μg/day which documents some absorption of aluminum from these compounds (13). Although the amount is small, if a dialyzed uremic patient was unable to excrete this small amount, this could appreciably alter body aluminum burden since total body aluminum is normally only about 30 mg.

In an unplanned controlled experiment, Flendrig et al (12) found that six patients, in a dialysis unit using dialysate prepared from water heavily contaminated with aluminum, developed the dialysis encephalopathy syndrome. In the same city but in other dialysis units, where water contamination was absent, no patient developed this syndrome. The aluminum concentration in brain and other tissues in patients dying of the encephalopathy syndrome was also much higher at this particular unit than that found in dialysis patients dying from other causes in the city's other dialysis units. Taken together this is additional evidence that the dialysis encephalopathy syndrome may be caused by aluminum intoxication and that under certain conditions the dialysate can be a source of the aluminum. Municipal water supplies commonly contain aluminum since aluminum is widely used as a flocculent in water treatment. A recent study (14) would suggest that virtually any aluminum present in the dialysate represents a potential source of aluminum uptake by the patient since aluminum is bound in plasma and readily transverses the dialyzing membrane. The plasma binding also prevents the removal of aluminum during dialysis when the patient is dialyzed with aluminum free dialysate. It can be concluded that aluminum deposited in the gray matter most likely comes from two sources, the aluminum contamination present in the dialysate and the orally administered phosphate binders. This would tend to explain the marked difference in incidence of this disease since centers using water high in aluminum concentration would be expected to have a higher incidence which would manifest itself earlier than centers using aluminum free dialysate.

MANAGEMENT

Therapy is largely symptomatic. Nadel & Wilson (15) feel that some of the symptoms and the EEG changes represent a form of seizure disorder. They were able to show that diazepam reversed the symptoms and normalized the EEG. They suggested that this drug might be effective for the chronic management of this state. However, I as well as other investigators (9) have found that although diazepam is effective early in the course of the disease, it becomes less effective later in the course of the illness, and it appears to have no effect on altering the inevitable fatal outcome of this illness. Since seizures frequently occur during or immediately following dialysis, diazepam, if given with the commencement of dialysis, is effective in preventing seizures in a number of patients.

Changing dialytic techniques either by increasing or decreasing frequency or time on dialysis has no lasting beneficial effects. In general, results with renal transplantation in patients with clinically apparent encephalopathy syndrome have been disappointing. In the majority of patients the disease has progressed in spite of good renal function. In addition, several patients have been described who initially developed the disease in the early transplant period. It seems likely that in these patients the disease was not recognized prior to transplantation and, as stated above, the surgical trauma caused the disease to become clinically apparent. Although the results with transplantation have not been encouraging, it still might be worthwhile transplanting patients early in the course of their disease, when only EEG abnormalities are present, to see if the disease might be arrested or reversed at this stage.

CONCLUSION

Increasing evidence has been presented which suggests that the dialysis associated encephalopathy syndrome is a distinct syndrome characterized by typically clinical and electroencephalographic features. This syndrome is widespread in distribution and a major cause of death in chronic dialysis patients. Although the specific etiology of this syndrome has not been clearly defined, aluminum intoxication appears to be the most likely possibility at this time.

Literature Cited

1. Alfrey, A. C., Mishell, J. M., Burks, J., Contiguglia, S. T., Rudolph, H., Lewin, E., Holmes, J. H. 1972. Syndrome of dyspraxia and multifocal seizures associated with chronic hemodialysis. *Trans. Am. Soc. Artif. Int. Organs* 18:257–61
2. Alvarez-Ude, F., Gentile, M., Ward, M. K., Appleton, D. R., Kerr, D. N. S. 1975. Residual symptoms in patients on regular haemodialysis. *Proc. Int. Congr. Nephrol., 6th,* pp. 650–56
3. Mahurkar, S. M., Dhar, S. K., Salta, R., Meyers, L. Jr., Smith, E. C., Dunea, G. 1973. Dialysis dementia. *Lancet* 2:1412–15
4. Chokroverty, S., Bruetman, M. E., Berger, V., Reyes, M. G. 1976. Progressive dialytic encephalopathy. *J. Neurol. Neurosurg. Psychiatry* 39:411–19
5. Schreeder, M. T. 1976 *Dialysis encephalopathy: An epidemiologic investigation.* Presented at Ann. Epidemic Intell. Serv. Conf., Atlanta, 1972
6. Ball, J. H., Butkus, D. E., Madison, D. S. 1977. Effect of subtotal parathyroidectomy on dialysis dementia. *Nephron* 18:151–56
7. Mahurkar, S. D., Kamath, C. R. V., Mamdani, B. H., Meyers, H. L. Jr., Smith, E. C., Dunea, G. 1974. Studies in dialysis dementia. *Am. Soc. Nephrol.* 7:57
8. Burks, J. S., Alfrey, A. C., Huddlestone, J., Norenberg, M. D., Lewin, E. 1976. A fatal encephalopathy in chronic haemodialysis patients. *Lancet* 1:764–68
9. Barratt, L. J., Lawrence, J. R. 1975. Dialysis-associated dementia. *Aust. NZ J. Med.* 5:62–65
10. Ward, M. K., Pierides, A. M., Fawcett, P., Shaw, D. A., Perry, R. H., Tomlinson, B. E., Kerr, D. N. S. 1976. Dialysis encephalopathy syndrome. *Proc. Eur. Dial. Transplant Assoc.* 13:348–54
11. Alfrey, A. C., LeGendre, G. R., Kaehny, W. D. 1976. The dialysis encephalopathy syndrome. Possible aluminum intoxication. *N. Engl. J. Med.* 294:184–88
12. Flendrig, J. A., Kruis, H., Das, H. A. 1976. Aluminum intoxication: the cause of dialysis dementia? *Proc. Eur. Dial. Transplant Assoc.* 13:355–61
13. Kaehny, W. D., Hegg, A. P., Alfrey, A. C. 1977. Gastrointestinal absorption of aluminum from aluminum containing antacids. *N. Engl. J. Med.* 296:1389
14. Kaehny, W. D., Alfrey, A. C., Holman, R. E., Shorr, W. J. 1977. Aluminum transfer during hemodialysis. *Kidney Int.* In press
15. Nadel, A. M., Wilson, W. P. 1976. Dialysis encephalopathy: A possible seizure disorder. *Neurology* 26:1130–34

Ann. Rev. Med. 1978. 29:99–116
Copyright © 1978 by Annual Reviews Inc. All rights reserved

DIGESTION AND ABSORPTION OF PROTEIN[1]

◆7270

Hugh J. Freeman, M.D.[2] *and Young S. Kim, M.D.*
Gastrointestinal Research Laboratory, Veterans Administration Hospital and
Department of Medicine, University of California School of Medicine, San
Francisco, California 94121

INTRODUCTION

Protein digestion and absorption may be altered in various pathological states with potential nutritional consequences (1–3). This review summarizes normal mechanisms of protein assimilation and details the alterations documented in various diseases.

The minimum daily requirement for digestible nutritionally important protein in a 70-kg man, necessary to maintain positive nitrogen balance, has been estimated to be about 45 g (4). In addition to this exogenous dietary source, a considerable amount of endogenous protein is available.

EXOGENOUS AND ENDOGENOUS SOURCES OF PROTEIN

Exogenous dietary protein totaling 70–100 g daily accounts for about 11–14% of the average caloric intake of North Americans (5). Various factors determine the nutritive value of food protein including quantity and quality of its constituent amino acids as well as its digestibility. Dietary protein must supply essential amino acids, i.e. amino acids whose carbon skeleton cannot be formed in man (Table 1). Nonessential amino acids, on the other hand, can be supplied from other sources since the carbon skeleton arises from intermediary products of carbohydrate and fat metabolism, and the amino group is obtained by transamination. Therefore, positive nitrogen balance and effective protein synthesis depend not only on an adequate supply of essential amino acids, but also on the speed and efficiency with which the nonessentials are made available. Not all

---

[1] Supported in part by US Public Health Service grant AM-17938 from the National Institute of Health and by the Veterans Administration Medical Research Service.

[2] H. J. F. is a recipient of a Medical Research Council of Canada Research Fellowship.

0066-4219/78/0401-0099$01.00

protein ingested is necessarily digestible to the same extent. Except for rare examples (i.e. egg albumin), most natural and processed foods contain protein complexed to fat or carbohydrate which may hinder efficient proteolysis. Furthermore, food protein damaged by heating during manufacture or by storage under unfavorable conditions may be highly resistant to enzymic digestion. This is possibly related to interactions between various functional groups leading to new amide or ester linkages within and between peptide chains that resist hydrolytic enzyme action or impede enzyme access to adjacent peptide bonds (6). With increasing complexity of food processing techniques, additional adverse nutritional effects may occur.

The major sources of endogenous protein include digestive gland secretions (at least 20–30 g daily), desquamated cells resulting from active replacement of the lining epithelium (at least 30 g daily), and finally, small amounts of plasma proteins (approximately 1–2 g daily) (7). Endogenous proteins found in secretions from salivary glands, stomach, intestine, biliary tract, and pancreas consist primarily of hydrolytic enzymes and glycoproteins or mucins. Desquamated cells resulting from active gastrointestinal mucosal renewal provide an additional source of endogenous protein. The precise quantities of protein derived from these desquamated cells in the normal gastrointestinal tract and after alterations by various disease states are unknown. However, in patients with celiac sprue, the replacement of lining epithelium is accelerated both in vitro (8) and in vivo (9, 10). While mucosal injury itself may partially account for shedding of injured surface epithelium and resultant protein loss, the increased turnover of epithelial cells results in greater epithelial cell desquamation and probably further contributes to protein loss in these patients. Plasma proteins in small, but significant, quantities reach the small intestinal lumen each day and contribute to the endogenous protein load in normal subjects. Albumin, gamma globulin, and other serum proteins have been demonstrated in the gastrointestinal contents of normal individuals by electrophoretic and immunologic techniques (11, 12). However, in some disease states, there is excessive loss of serum protein into the gastrointestinal tract resulting in hypoproteinemia, i.e. protein-losing gastroenteropathy (13). Regardless of their origin these endogenous proteins are likely to be extensively hydrolyzed and reused as a

Table 1 Essential amino acids

Leucine	Valine
Isoleucine	Methionine
Lysine	Cystine[a]
Tryptophan	Phenylalanine
Threonine	Tyrosine[b]
Arginine[c]	

[a] Derived from methionine.
[b] Derived from phenylalanine.
[c] Children only (5).

nitrogen source by the host. Since the healthy adult excretes only 1–2 g of fecal nitrogen per day (equivalent to 6–12 g of protein), the processes involved in digestion and absorption of both exogenous dietary (70–100 g per day) and endogenous (35–200 g per day) protein appear to be very efficient.

NORMAL PROTEIN DIGESTION AND ABSORPTION

In this review, protein digestion and absorption is divided (for descriptive purposes) into gastric, pancreatic, and intestinal phases. After a meal, however, these phases occur in concert influenced by such factors as gastric emptying and rate of meal transit through the intestine.

Gastric Phase

Protein digestion appears to be first initiated in the stomach. The major gastric protease, pepsin, is activated from a zymogen precursor, pepsinogen, by selective enzymic cleavage of a small basic peptide (14–16). Pepsin itself activates pepsinogen, so that the process becomes autocatalytic and functions only in a confined acidic pH range (16, 17). The major peptic digestion products are polypeptides with N-terminal amino acids including phenylalanine and leucine as well as small amounts of oligopeptides and amino acids (18). Electrophoretically, seven distinct pepsinogens have been identified in gastric mucosa and are divided into two groups. Group 1 pepsinogens (pepsinogens 1 to 5) are limited to fundic gland mucosa and are found in serum and urine. Group 2 pepsinogens (pepsinogens 6 and 7) are found in fundic, antral, pyloric, and duodenal mucosa as well as in serum, but not in urine (19). Each pepsinogen is thought to give rise to a distinct pepsin enzyme (16).

The gastric phase of protein digestion is influenced by hormonal factors (i.e. gastrin, cholecystokinin, and secretin) as well as nonhormonal factors (i.e. gastric acidity, mixing, and the rate of gastric emptying). Hormone release may be stimulated by protein, polypeptides, or amino acids which leads to secretion of acid and pepsin. For example, gastrin stimulates both acid and pepsin secretion (20) while secretin, which inhibits basal- and gastrin-stimulated acid secretion, is a strong stimulant of pepsin secretion (21). The rate of gastric emptying, and subsequent transfer of gastric contents into the small intestine, appears to be retarded by the activity of receptors not only sensitive to duodenal acid, fat, and osmotic pressure, but also to amino acids (22). Earlier studies suggested that gastric emptying was related to the osmotic pressure exerted by each amino aicd (23). Subsequently, at least one amino acid, *l*-tryptophan, was demonstrated to delay emptying under physiological conditions (24). Finally, the pyloric sphincter, a high pressure zone which relaxes with peristalsis, contracts in response to the presence of luminal amino acids (possibly hormonally mediated) (25). These studies indicate that protein digestion products influence various physiological processes in the stomach.

It would appear from available data that the gastric phase plays a minimal, and possibly dispensible, role in overall protein assimilation. Some patients with total gastrectomy can assimilate dietary protein adequately and remain in positive nitrogen balance (26). Furthermore, patients with pernicious anemia and little or no gastric proteolytic activity can also handle dietary protein adequately (5). While pepsin may not be considered directly essential in protein digestion, its principle function may be indirect. Pepsin may unmask amino acids or peptides within ingested whole-protein molecules which results in a release of pancreatic enzymes.

Pancreatic Phase

The pancreatic phase, or alternatively, the intraluminal phase, is felt to be more critical than the gastric phase. Peptic digestion products are more potent secretagogues for endocrine cells in the small-intestinal mucosa than whole protein (18) (Figure 1). The endocrine cells secrete cholecystokinin and secretin which cause pancreatic acinar-cell release of hydrolytic enzyme precursors (18, 27). The hormones have an additional action on the mucosal cells which causes a release of the brush-border glycoprotein enzymes; alkaline phophatase, sucrase, and enteropeptidase (enterokinase) (28).

Enteropeptidase converts trypsinogen to trypsin (15, 29, 30) by scission of a hexapeptide from the amino-terminus through hydrolysis of the peptide linkage between lysine-6 and isoleucine-7 which releases a hexapeptide and trypsin. The resultant trypsin in turn activates trypsinogen itself and other pancreatic zymogens secreted into the duodenum (30). Activated pure pancreatic juice contains both endopeptidases and exopeptidases. Endopeptidases such as trypsin, chymotrypsin, and elastase each internally cleave protein or polypeptide substrates at different specific peptide linkages. Exopeptidases consist primarily of carboxypeptidase (A and B) which act on terminal peptide bonds at the carboxyl end of protein. A trace of aminopeptidase activity can also be detected in pancreatic juice (15, 31). The products of the combined action of these enzymes are small peptides and amino acids. These small peptides consist of 2–6 amino acid residues and account for over 60% of luminal α-amino nitrogen; the remaining 40% is associated with amino acids (32, 33). Contrary to earlier reports

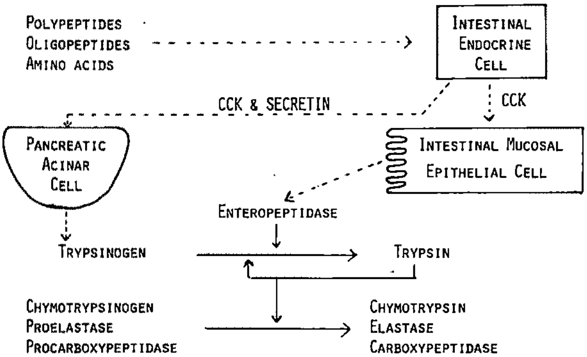


Figure 1 Secretion and activation of pancreatic enzymes.

of rapid and almost complete protein digestion and absorption in upper jejunum, recent evidence indicates that a considerable amount of protein may be present in the distal small intestine even hours after ingestion of a test meal (34). However, quantitative estimates of the amount of dietary protein and various protein digestion products found in different regions of the small intestine during the digestive process are not yet available.

Intestinal Phase

Following oral protein administration, the amino nitrogen in postprandial blood is almost entirely in the form of free amino acids (35). A few small peptides, however, do appear in portal blood. These include proline- and hydroxyproline-containing peptides after gelatin ingestion (36, 37), and both carnosine and anserine after ingestion of certain meats, specifically chicken breast (38).

Significant amounts of whole protein and other large macromolecules are absorbed intact by fetal and neonatal small intestine (39, 40). The serum of newborns contains bovine serum albumin after oral feeding (41) and, compared to the serum of adults, a higher ratio of antibodies to food antigens (42, 43). Immunoglobulin from colostrum appears also to be absorbed intact by the neonate which affords passive immunity for the newborn (40). This process of intact transport of significant amounts of protein appears confined to the first 48 hr after birth. Studies in mature mammalian gut with ferritin (44) and horseradish peroxidase (45) indicate that uptake of small amounts of macromolecules occurs intact. While indirect, these studies suggest that the normal human adult intestine is permeable to macromolecules, perhaps not in sufficient quantities to be of nutritional importance, but in quantities that may be active antigenically or biologically (40). In human adults, several studies (46–48) suggest that macromolecules can cross the mature mucosal barrier under normal physiological conditions. For example, milk antibodies are found in the serum after milk protein is ingested orally (48).

The final steps in digestion of most luminal peptides appear to be associated with the intestinal mucosal cells. High aminopeptidase activity is found in the intestinal mucosal cells, and only a trace of activity is detected in jejunal contents (49). This peptidase activity is located in two main subcellular fractions: the cytoplasmic membrane and the brush-border membrane (50). Although both are aminopeptidases, striking differences in electrophoretic mobilities, physicochemical properties, and substrate specificities are found (51). For example, the proportion of enzyme activity in these two subcellular fractions varies considerably depending upon the type and chain length of substrate. The brush-border membrane fraction is associated with less than 10% of the total cellular activity for dipeptides but as much as 60% for tripeptides (50, 52). With a tetrapeptide, almost all of the activity is found in the brush-border fraction (53). Furthermore, proline-containing peptides appear to be hydrolyzed almost exclusively by cytoplasmic peptidases, whereas brush-border peptidases have almost no activity against these peptides (1). These results indicate that the enzymes located in the two subscellular fractions are distinct.

As mentioned, the major products of intraluminal digestion of protein are amino acids and small peptides. Amino acid transport in the small intestine has been extensively reviewed (3, 54, 55) and appears to involve several specific carrier-mediated active transport systems coupled to sodium transport (56, 57). The occurrence of more than one distinct amino acid transport system was indicated by studies both in man and in experimental animals. Studies in subjects with some rare inherited human small-intestinal amino acid transport defects also support this conclusion (58–60). These transport systems include:

1. neutral (monoaminomonocarboxylic) amino acids: alanine, glycine, serine, threonine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, asparagine, histidine, cysteine, methionine, and citrulline;
2. dibasic (diamino) amino acids: lysine, arginine, ornithine, and the neutral amino acid, cystine;
3. dicarboxylic (acidic) amino acids: glutamic and aspartic acids;
4. glycine and imino acids: glycine, proline, and hydroxyproline.

Some amino acids have affinity for more than one carrier system, exemplified by glycine. Studies in man (61–63) indicate that amino acid absorption rates are higher in proximal jejunum than ileum.

Previously it was believed that complete hydrolysis of protein occurred in the intestinal lumen followed by uptake of free amino acids alone. However, the major intraluminal digestion products include small peptides. The following summarizes data from various laboratories (1, 2, 64, 65) on peptide transport:

1. Mucosal uptakes of peptides and amino acids are independent processes.
2. Dipeptides and tripeptides are actively transported against a concentration gradient.
3. Dipeptides and tripeptides appear to share a common transport mechanism.
4. Rates of amino acid absorption are frequently faster from peptides than from free amino acids.
5. Whereas amino acids are absorbed more rapidly in the proximal intestine, peptides appear to be absorbed well in both proximal and distal small intestine.

A schematic representation of possible mechanisms for peptide hydrolysis and transport is illustrated in Figure 2. There are at least two general mechanisms for transport of small peptides. First, small peptides poorly hydrolyzed by brush-border membrane peptidases appear to be transported intact across the brush-border membrane, thus bypassing the brush-border peptidases, and are later hydrolyzed by the cytoplasmic peptidases. As has been mentioned, dipeptides and tripeptides appear to be transported by common mechanisms distinct from those of amino acids. Second, small peptides having a higher affinity for brush-border membrane peptidases likely undergo hydrolysis primarily at the brush-border membrane. The hydrolysis products, including free amino acids and dipeptides or tripeptides, are then absorbed by their respective transport mechanisms. For longer peptides, the major, if not exclusive, site of hydrolysis appears to be the brush-border membrane, with subsequent absorption of the products of peptide hydrolysis including free amino acids, dipeptides, and tripeptides.

DEVELOPMENTAL ASPECTS

Studies in the developing human fetus and neonate have improved our understanding of the fate of ingested protein, its role in maintaining fetal homeostasis, and its adaptation to neonatal life.

Gastric Phase

Peptic activity is detected in the human fetal stomach during the 16th wk of gestation and increases markedly between the 28th and 40th wk (66, 67). However, a "fetal pepsinogen" (68) has also been demonstrated immunochemically in fetal gastric mucosa. This is of interest particularly in relation to oncodevelopmental aspects of enzymes in gastric malignancy, i.e. fetal pepsinogen may also be detected in gastric carcinoma, regardless of tumor location (68, 69). The relationship of peptic activity and gastric acid secretion in the fetal stomach as well as its role in fetal nutrition has not been well studied. Pepsin and acid appear to be secreted in parallel in the neonatal period and by two years of age pepsin secretion has reached adult levels. While some have suggested that this low pepsin and acid content in the stomach may have detrimental nutritional effects in the premature infant, no firm evidence is available to support this contention (70).

Pancreatic Phase

Zymogen granules are detected in the fetal pancreas by the fifth month of gestation (71, 72). Tryptic activity, however, may be present as early as 16 wk and increases with advancing gestational age, especially after the 28th wk (66, 73). Cholecystokinin and secretin administration produce stimulation of exocrine secretion in the neonate (74). When timed output and maximum stim-

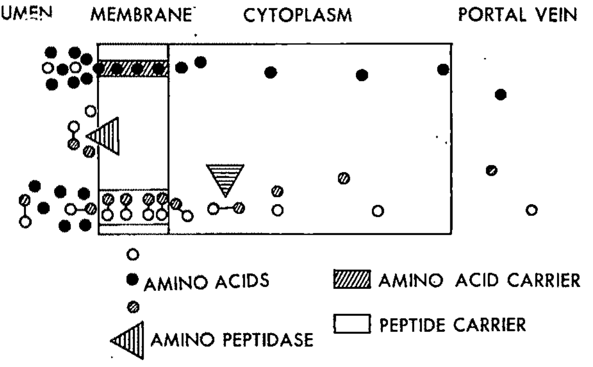


Figure 2 Schematic representation of intestinal transport of amino acids and peptides.

ulation of pancreatic trypsin activity are examined, they are low in the first 24 hr after birth, somewhat increased by 1 wk, greatly increased by one month, and markedly increased by nine months (70).

Intestinal Phase

Although absorption of intact protein has been demonstrated for both fetal and neonatal intestine, little is known about digestion and utilization of protein in utero. Amino acid and peptide transport in the human fetus has not been well studied although transport of *l*-alanine has been demonstrated (75).

More information is available on human small-intestinal enzyme ontogenesis. Leucine aminopeptidase was demonstrated histochemically in the 7–8 wk fetus which suggests that enzymatic differentiation is far advanced relative to other aspects of protein digestion and absorption (76, 77). Specific activity of leucine aminopeptidase is higher in proximal small intestine early in fetal life but, with increasing gestational age, this enzyme increases more distally and causes a significant proximal-to-distal gradient—the distal values being at least twice as great after 16 wk gestation (78, 79).

Finally, recent investigations suggest that the fetal colon, because of its similarity to small intestine, may have an absorptive mechanism in utero (80). Further studies are required to examine the potential role of the colon in fetal digestion and absorption of proteins.

DERANGEMENTS IN PROTEIN DIGESTION AND ABSORPTION

Protein digestion, absorption, or both may be impaired in certain disease states. Furthermore, defects in different phases of digestion and absorption may occur in the same disorder. After gastric resection, for example, all phases may be altered leading to significant malnutrition in the patient (81).

Gastric Phase

Disturbances in the gastric phase may alter the fate of orally ingested protein. Significant weight loss, postoperative malnutrition, and nitrogen malabsorption may occur following total and subtotal gastrectomy. In part, this is due to reduced intake related to the absence or limited size of the gastric reservoir (82–84), reduced peptic digestion in the residual stomach (85), and rapid intestinal transit of ingested food, which results in poor mixing of digestive pancreatic secretion with food (86, 87). Few investigations, however, have been undertaken to explain these theoretical considerations. Low trypsin concentrations in the proximal gut are found following subtotal gastrectomy during the initial hour after ingestion of a liquid test meal (88–90). However, the capacity of the pancreas for enzyme secretion is not significantly altered by gastric surgery. Rather, pancreatocibal asynchrony or impaired duodenal stimulation with reduced postprandial serum levels of cholecystokinin have been observed following hemigastrectomy and gastrojejunostomy (91). Newly developed perfusion methods (92, 93) allow the study of pancreatic enzyme output following ingestion of liquid (94) and solid (95) test meals. In these studies (94, 95), low intraluminal trypsin concentrations were found in patients with subtotal gastrectomy and gastroduodenostomy (Billroth 1) or gastrojejunostomy (Billroth 2), as well as in patients with truncal vagotomy with pyloroplasty, although total amount of trypsin secreted was unchanged. Rapid gastric emptying of liquid test meals with subsequent dilution of intraluminal content appeared responsible. Patients with vagotomy tend to have less postoperative weight loss and nutritional disturbance than patients with subtotal gastrectomy. This may be related to slower gastric emptying of solid meals in the vagotomy-pyloroplasty group (95) which allows reduced passage of chyme into the small intestine and subsequently greater efficiency of digestion and absorption. After truncal vagotomy, a reduction in both basal peptic activity (96) and pancreatic output of trypsin (89) has been observed. However, the effect on overall protein assimilation has not been assessed following truncal or newer, limited and more selective, forms of vagotomy.

Defective protein assimilation on the basis of a primary deficiency of pepsinogen or pepsin has not been described. Reduced or absent peptic activity has been observed (97) in different variants of pernicious anemia with and without achlordydria; included were patients with and without gastric mucosal inflammatory change, mucosal "atrophy," and deficient or normal intrinsic factor secretion (Imerslund's syndrome). In these patients no evidence of malnutrition was observed. Secondary deficiency of pepsin or its precursors may occur in various gastric mucosal inflammatory disorders, particularly if they involve the stomach severely and diffusely.

Pancreatic Phase

The pancreatic phase is more critical than the gastric phase to effective protein assimilation in man. Single specific or generalized enzymatic deficiencies occur but, with the exception of enterokinase and trypsinogen-trypsin deficiencies, single enzymatic deficiencies are not clinically significant (31).

Intestinal enterokinase deficiency is a rare disorder characterized by vomiting, diarrhea (often from birth), hypoproteinemia, anemia, failure to thrive, and edema (98, 99). Most patients are seen before eight years of age. In some, the disorder may be inherited, since it is described in siblings. The most important laboratory tests suggesting the disorder are normal pancreatic amylase and lipase activity associated with absent duodenal proteolytic enzymes. When hypoproteinemia is severe in the early phases, amylase and lipase activity may be reduced and steatorrhea may be present. These conditions return to normal with adequate therapy. Proteolytic enzymes, however, remain permanently reduced. The zymogram activation test (31) confirms enterokinase deficiency. Trypsin, chymotrypsin, and carboxypeptidase activities are measured before and after addition of enterokinase to a sample of duodenal juice. With enterokinase deficiency, a marked increase in activity of these three enzymes occurs. Patients with pancreatic or small-intestinal disorders (i.e. celiac sprue) appear always to have a negative test (31,100). Therefore, the test may be specific and should be performed whenever the disorder is suspected. Enzyme measurement in a small-intestinal biopsy is also useful. No other structural or functional abnormality has been detected. Therapy involves administration of pancreatic extract (31). Whether this replacement causes activation of the patient's own zymogens, or whether therapeutic effect is due to simple replacement of missing enzymes, is unknown. At present, therapeutic trials with a porcine purified enterokinase preparation are in progress (31). Amino acid mixtures such as Nutramingen®, absorbed in the absence of proteolytic pancreatic enzymes, have been used as a growth supplement in these patients (31). Variability in the response to pancreatic extracts occurs, but the mechanism is unknown. Enterokinase deficiency also occurs secondarily to intestinal mucosal disease, but its activity is not reduced to the same extent as that of disaccharidases. This may be because enterokinase also occurs in relatively young and less differentiated enterocytes (31, 101). Usually the reduction in enterokinase activity parallels the severity of the small-intestinal mucosal lesion. However, celiac sprue may be an exception to this rule.

Isolated trypsinogen-trypsin deficiency presents a clinical picture similar to enterokinase deficiency (102–104). Addition of trypsin to duodenal juice activates other pancreatic proenzymes—chymotrypsinogen and procarboxypeptidase. In addition, disturbed activation of trypsinogen has been associated with biliary atresia (105). Again, the mechanism causing abnormally low levels of trypsinogen-trypsin is unknown.

Generalized enzyme deficiency may result from exocrine pancreatic insufficiency. This may be associated with gastric surgery or caused directly by reduced functioning of the exocrine pancreatic mass. The pathophysiologic basis for reduced pancreatic enzyme activity following gastric surgery has been discussed in the preceding section. The reserve capacity of the pancreas is considerable, and significant nitrogen loss occurs only when trypsin output is less than 10% of normal output (106). In other words, over 90% of the exocrine pancreas must be destroyed before significant maldigestion occurs. This serves to explain the frequent observation that little detectable maldigestion occurs until very late in the course of chronic pancreatitis (107).

For evaluation of impaired pancreatic function, the analysis of duodenal contents following hormonal stimulation or a test meal are useful (108, 109). This requires duodenal intubation under fluoroscopic guidance with care taken to prevent contamination by gastric content. Enzymatic activity (including trypsin) and electrolytes are generally measured before and after pancreatic stimulation by cholecystokinin and secretin (109). In patients with more advanced pancreatic disease, feeding of a liquid test meal (108) with measurement of duodenal trypsin levels provides a relatively simple test of pancreatic function. A similar test using a meal of essential amino acids has been advocated to provide as great a response as direct stimulation with cholecystokinin (110). The examination of the digestion products in serum, intestinal content, and urine, following oral ingestion of gelatin or casein, has been shown to be useful in assessing exocrine pancreatic function in a few studies (111, 112). However, its clinical usefulness needs to be established in a large clinical series. A favorable reduction in azotorrhea with oral pancreatic enzyme therapy may also be a useful clinical parameter of exocrine pancreatic function.

The use of pancreatic enzymes in pancreatic insufficiency has been reviewed (113, 114). Briefly, the goals of therapy are to restore digestion and absorption to normal as well as to maintain positive caloric and nitrogen balance with tolerable amounts of enzyme. Much of the literature has dealt with available preparations, their comparison, relative efficiency and potency, frequency and dose administered, and the end point of therapy, particularly in relation to reduction of steatorrhea. However, oral enzyme replacement therapy may not reduce steatorrhea to the same extent as it reduces azotorrhea. In assessing each parameter, an important consideration should be the effect of enzyme replacement on actual digestion and absorption of protein. Unfortunately, as indicated previously, this important information is still forthcoming.

Intestinal Phase

In contrast to the well-known primary disaccharidase deficiency states associated with disaccharide malabsorption, no analogous primary peptidase syndromes have been recognized. Frazer (115–117) initially suggested that celiac sprue might be due to a deficiency of a peptide-splitting enzyme in the small-intestinal mucosa. Subsequent studies supported (118–121) or refuted (122–125) his original hypothesis. In general, low peptidase activity is observed in untreated celiac sprue and, with a gluten-free diet, returns to normal (126). This suggests that reduced peptidase activities in celiac sprue may be secondary to mucosal damage. Similar findings are observed in abnormal mucosa from patients with dermatitis herpetiformis (127), malnutrition (128), Crohn's disease (129), folic acid deficiency (130), and vitamin B_{12} deficiency (130) as well as in patients with "flat" mucosa who fail to respond clinically to gluten restriction (126). However, one cannot exclude the possibility of primary peptidase deficiency in celiac sprue unless all peptide substrates derived from the "toxic" fraction of gluten are examined. Generalizations based upon total mucosal peptidase activity may not be completely valid since subcellular fractionation into brush-border and cytosol components indicates these separate families of enzymes are not altered in parallel. For example, after starvation, marked reduction in brush-border peptidase activity occurs while cytoplasmic activity increases significantly (131). With purification of cytoplasmic and brush-border peptidases as well as characterization of their substrate specificity and kinetic properties, the significance of observations on clinical material will become more apparent. The clinical and nutritional significance of possible altered peptide hydrolase activity in various human disease states remains to be defined.

Most primary amino acid transport defects (Table 2) are very rare and were initially described as renal tubular transport disorders. Only later was deranged small-intestinal transport appreciated (59, 60).

The basic defect in Hartnup disease is reduced proximal renal tubular and small-intestinal transport of the neutral amino acids (55). The intestinal absorptive defect results in diminished blood tryptophan rise following oral loading with this amino acid as well as increased fecal excretion of tryptophan or tryptophan metabolites. Intestinal absorption of metabolites produced by bacterial degradation of unabsorbed tryptophan may contribute to the transient neurologic and psychiatric symptoms as well as other abnormal biochemical findings. For example, dietary tryptophan is an important precursor for endogenous nicotinamide. Impaired absorption of this amino acid may cause the pellagra-like skin rash and neuropsychiatric findings known to be improved by dietary supplementation of nicotinamide. Despite continued nitrogen loss in these patients, associated with impaired free amino acid uptake, severe nutritional imbalance is uncommon. This may be explained by our current understanding of separate transport mechanisms for amino acids and peptides. In this disorder, it has recently been observed (132) that transport of peptides is unaffected while neutral amino acid transport is defective.

Cystinuria is characterized as a defect in renal tubular and small-intestinal absorption of cystine, ornithine, arginine, and lysine (55). Because cystine is relatively insoluble in acid urine, recurrent urinary tract cystine stone formation

Table 2 Specific intestinal amino acid transport disorders

Defect	Substrate(s)	Clinical features
Hartnup disease	Neutral amino acids	Pellagra-like rash, cerebellar ataxia, neuropsychiatric disorder, neutral aminoaciduria
Blue diaper syndrome (tryptophan malabsorption)	Tryptophan	Indigo blue discoloration of diapers, no tryptophanuria, growth retardation, hypercalcemia.
Oasthouse urine disease (methionine malabsorption)	Methionine	Seizures, mental retardation, episodic hyperpnea, diarrhea, urinary odor with increased α-hydroxybutryric acid.
Cystinuria	Cystine, ornithine, arginine, lysine	Recurrent urinary tract calculi, hereditary pancreatitis, cerebral dysfunction, cystine-dibasic aminoaciduria.
Hyperdibasic-aminoaciduria	Lysine	Familial protein intolerance diarrhea, vomiting, failure to thrive, hepatosplenomegaly, dibasic aminoaciduria (ornithine, arginine, lysine)
Joseph's syndrome (iminoglycinuria)	Proline, hyproxyproline, glycine	No reproducible clinical picture, aminoaciduria (imino acids, glycine)

is a significant complication. The stones are radiopaque due to high sulfur content. Cystinuria is traditionally diagnosed by finding flat hexagonal crystals in the urine and a positive cyanide nitroprusside test. If this test is positive, high-voltage urinary electrophoresis will confirm the diagnosis. Homocystinuria, the only other disorder giving a positive cyanide nitroprusside test, can readily be distinguished clinically and excluded definitively by urinary electrophoresis (55). Evidence for a similar uptake defect in the small intestine includes a subnormal blood cystine and arginine rise after orally ingesting these amino acids as well as increased fecal levels. The uptake defect has also been demonstrated in small-intestinal mucosal biopsies. As in Hartnup disease, absorption of affected amino acids is normal or much improved when given as small peptides compared to administration as free amino acids (133).

An isolated defect in renal and intestinal transport termed hyperdibasicaminoaciduria has also been described (134, 135). This disorder may be distinguished clinically (Table 2) and chemically from classical cystinuria. At least two clinically significant intestinal transport disorders, tryptophan (136) and methionine malabsorption (137, 138), have been described without consistently detectable loss of these amino acids in the urine. Finally, iminoglycinuria (139) is rarely recognized, possibly because of failure to manifest a consistent clinical syndrome.

A generalized inhibition of amino acid transport is exemplified by the oculocerebralrenal syndrome of Lowe (55). This is an X-linked recessive defect characterized by glaucoma, hypotonia, mental retardation, and a generalized aminoaciduria. Although peptiduria may be seen, no analogous defects in primary intestinal peptide uptake have been discovered.

Malabsorption of various substances including amino acids occurs in association with many small-intestinal disorders. In some disorders, this appears to be secondary to the mucosal abnormality. In patients with the mucosal lesion of celiac sprue, jejunal absorption of free amino acids (61, 140, 141) is reduced while jejunal dipeptide absorption is less severely affected (140, 141). The more severe proximal impairment of amino acid absorption (61) parallels the more severe proximal mucosal disease in celiac sprue. Similarly reduced absorption could not be documented in dermatitis herpetiformis (140). This may reflect less severe and less extensive mucosal involvement. In tropical sprue amino acid uptake from small peptides is spared relative to its constituent free amino acids (142). Peptide transport is also more resistant to various nutritional and physiological alterations than is amino acid transport. In protein-calorie malnutrition following jejunoileostomy for obesity, a reduction in jejunal absorption of the free amino acid leucine occurs without any alteration in the absorption of the dipeptide, glycylleucine (143). Acute bacterial infections in man increase jejunal absorption of glycine (144) (possibly as a result of semistarvation), while absorption of glycylglycine is not altered (145). In view of the relative sparing of peptide transport processes when compared to amino acids in various pathological states thus far examined, a possible role for peptides as an early nutritional supportive measure is evident. The use of synthetic free amino acids as a nitrogen source has been previously questioned (7). They are hyperosmotic, more costly, and less palatable than comparable synthetic peptide diets. Recently, absorption of a synthetic peptide diet was compared to absorption of a synthetic amino acid mixture in human subjects in our laboratory (146). This study showed that small peptides were much less osmotically active, were absorbed as efficiently, and were more palatable than amino acids. The peptide diets, therefore, might be used effectively in patients with severe nutritional deficits secondary to maldigestion or malabsorption.

In small-intestinal bacterial overgrowth or stasis syndrome, significant evidence of protein wasting may be present, often severe enough to resemble protein-calorie malnutrition (147). In addition, reduced serum albumin concentrations and excessive fecal nitrogen excretion occurs. The pathogenesis for these phenomena is not clear, but it appears related either to conversion of dietary or endogenous amino acids into unusable forms by bacteria, or to enteric loss of protein. Organisms in the small intestine may elaborate enzymes that deaminate or decarboxylate amino acids. A patient with blind loop syndrome with marked hypoproteinemia was observed to have reduced albumin and fibrinogen synthesis associated with increased urea synthesis (148). This was thought to result from bacterial deamination of dietary protein with formation of ammonia and subsequent utilization for urea synthesis. Enteric loss of protein occurs in patients with intestinal stricture and inflamed small-intestinal mucosa. Although some morphological alterations may occur in association with bacterial overgrowth, it has not been determined whether bacterial overgrowth per se can cause enteric protein loss. Furthermore, uptake of *l*-leucine is impaired in experimental animals with self-filling blind loops which indicates that intestinal mucosal transport of amino acid may be affected in this condition (149). Several microbes elaborate tryptophanase which converts tryptophan to indole derivatives. These may be absorbed and metabolized to indicans and subsequently excreted. Following intake of oral antibiotics, urinary tryptophan metabolites decrease and stool tryptophan increases. This has also been observed in Hartnup disease which underlines the importance of intestinal microflora in this disorder. However, urinary indican excretion may also increase in the urine of patients with small-intestinal mucosal diseases. This appears to be related to delivery of increased amounts of unabsorbed tryptophan to the colon where indole compounds are formed prior to being absorbed and excreted in the urine.

Miscellaneous drug-induced absorptive defects after intake of neomycin, biguanides, ethacrynic acid, probenecid, phenylbutazone, and salicylates are recorded (150), but the precise mechanisms for altered absorption are uncertain.

FUTURE DIRECTIONS

Clearly, further study of protein digestion and absorption is indicated in normal subjects as well as in patients with various diseases and nutritional disorders. The therapeutic efficacy of peptide diets also requires further investigation in these patients.

Literature Cited

1. Kim, Y. S., Nicholson, J. A., Curtis, K. J. 1974. *Med. Clin. North Am.* 58:1397–1412
2. Matthews, D. M., Adibi, S. A. 1976. *Gastroenterology* 71:151–61
3. Their, S. O., Alpers, D. H. 1969. *Am. J. Dis. Child.* 117:13–23
4. Hellier, M. D., Holdsworth, C. D. 1975. In *Intestinal Absorption in Man,* ed. I. McColl, G. E. Sladen, pp. 143–86. London: Academic. 363 pp.
5. Albanese, A. A., Orto, L. A. 1973. In *Modern Nutrition in Health and Disease,* ed. R. S. Goodhart, M. E. Shils, pp. 28–88. Philadelphia: Lea & Febiger. 1153 pp.
6. Ford, J. E. 1975. In *Peptide Transport in Protein Nutrition,* ed. D. M. Matthews, J. W. Payne, pp. 183–203. Amsterdam: North-Holland. 503 pp.
7. Matthews, D. M. 1971. *J. Clin. Pathol.* 24: Suppl., *Roy Coll. Path.* 5:29–40
8. Trier, J. S., Browning, T. H. 1970. *N. Engl. J. Med.* 283:1245–50
9. Wright, N., Watson, A., Morley, A., Appleton, D., Marks, J., Douglas, A. 1973. *Gut* 14:603–6
10. Wright, N., Watson, A., Morley, A., Appleton, D., Marks, J. 1973. *Gut* 14:701–10
11. Gullberg, R., Olhagen, B. 1959. *Nature* 184:1848–49
12. Holman, H., Nickel, W. F. Jr., Sleisenger, M. H. 1959. *Am. J. Med.* 27:963–75
13. Waldmann, T. A. 1976. In *Gastroenterology,* ed. H. L. Bockus, Vol. 2, pp. 361–85. Philadelphia: Saunders. 1171 pp. 3rd ed.
14. Neurath, H., Walsh, K. A. 1976. *Proc. Natl. Acad. Sci. USA* 73:3825–32
15. Keil, B. 1971. In *The Enzymes,* ed. P. D. Boyer, Vol. 3, pp. 249–75 New York: Academic. 886 pp.
16. Rudick, J., Janowitz, H. D. 1974. In *Gastroenterology,* ed. H. L. Bockus, Vol. 1, pp. 405–18. Philadelphia: Saunders. 1148 pp. 3rd ed.
17. Piper, D. W., Fenton, B. 1965. *Gut* 6:506–8
18. Meyer, J. H., Kelly, G. A. 1977. *Am. J. Physiol.* 231:682–91
19. Samloff, I. M. 1971. *Gastroenterology* 60:586–604
20. Walsh, J. H., Grossman, M. I. 1975. *N. Engl. J. Med.* 292:1324–34, 1377–84
21. Rayford, P. L., Miller, T. A., Thompson, J. C. 1976. *N. Engl. J. Med.* 294:1093–1101, 1157–64
22. Cooke, A. R. 1975. *Gastroenterology* 68:804–16
23. Cooke, A. R., Moulang, J. 1972. *Gastroenterology* 62:528–32
24. Stephens, J. R., Cooke, A. R. 1974. *Clin. Res.* 22:606A (Abstr.)
25. Fisher, R., Cohen, S. 1973. *Gastroenterology* 64:67–75
26. Davenport, H. W. 1971. In *Physiology of the Digestive Tract,* pp. 191–96. Chicago: Yearbook. 229 pp.
27. Curtis, K. J., Kim, Y. S. 1975. *Gastroenterology* 68:1071 (Abstr.)
28. Götze, H., Adelson, J. B., Hadorn, B., Portmann, R., Troesch, V. 1972. *Gut* 13:471–76
29. Maroux, S., Baratti, J., Desneulle, P. 1971. *J. Biol. Chem.* 246:5031–39
30. Neurath, H., Dixon, G. H. 1957. *Fed. Proc.* 16:791
31. Hadorn, B. 1974. *Med. Clin. North Am.* 58:1319–31
32. Chen, M. L., Rogers, Q. R., Harper, A. E. 1962. *J. Nutr.* 76:235–41
33. Nixon, S. E., Mawer, G. E. 1970. *Br. J. Nutr.* 24:241–58
34. Adibi, S. A., Mercer, D. W. 1973. *J. Clin. Invest.* 52:1586–94
35. Dent, C. E., Schilling, J. A. 1949. *Biochem. J.* 44:318–35
36. Prockop, D. J., Sjoerdsma, A. 1961. *J. Clin. Invest.* 40:843–49
37. Bronstein, H. D., Haeffner, L. J., Kowlessar, O. D. 1966. *Gastroenterology* 50:621–30
38. Perry, T. L., Hansen, S., Tischler, R., Bunting, R., Berry, K. 1967. *N. Engl. J. Med.* 277:1219–27
39. Lev, R., Orlic, D. 1973. *Gastroenterology* 65:60–68
40. Walker, W. A., Isselbacher, K. J. 1974. *Gastroenterology* 67:531–50
41. Rothberg, R. M. 1969. *J. Pediatr.* 75:391–99
42. Lippard, V. W., Schloss, O. M., Johnson, P. A. 1936. *Am. J. Dis. Child.* 51:562–74
43. Stanfield, J. P. 1959. *Acta Paediatr. Scand.* 48:85–98
44. Bockman, D. E., Winborn, W. B. 1966. *Anat. Rec.* 155:603–22
45. Cornell, R., Walker, W. A., Isselbacher, K. J. 1971. *Lab. Invest.* 25:42–48

46. Wilson, S. J., Walzer, N. 1935. *Am. J. Dis. Child.* 50:49–57
47. Grusky, F. L., Cooke, R. E. 1955. *Pediatrics* 16:763–68
48. Korenblat, R. E., Rothberg, R. M., Minden, P. 1968. *J. Allergy* 41:226–35
49. Lindberg, T. 1966. *Acta Physiol. Scand.* 69:Suppl.285
50. Kim, Y. S., Birtwhistle, W., Kim, Y. W. 1972. *J. Clin. Invest.* 51:1419–30
51. Kim, Y. S., Kim, Y. W., Sleisenger, M. H. 1974. *Biochim. Biophys. Acta* 370:283–96
52. Peters, T. J. 1970. *Biochem. J.* 120:195–203
53. Chung, Y. C., Silk, D. B., Kim, Y. S. 1976. *Clin. Res.* 24:282A (Abstr.)
54. Saunders, S. J., Isselbacher, K. J. 1966. *Gastroenterology* 50:586–95
55. Schriver, C. R., Rosenberg, L. E. 1973. *Amino Acid Metabolism and Its Disorders*. Philadelphia: Saunders. 491 pp.
56. Gray, G. M., Cooper, H. L. 1971. *Gastroenterology* 61:535–44
57. Schultz, S. G., Curran, P. F. 1970. *Physiol. Rev.* 50:637–718
58. Baron, D. N., Dent, C. E., Harris, H., Hart, E. W., Jepson, J. B. 1956. *Lancet* 2:421–28
59. Milne, M. D., Crawford, M. A., Girao, C. B., Loughridge, L. W. 1960. *Q. J. Med.* 29:407–21
60. Milne, M. D., Asatoor, A. M., Edwards, K. D., Loughridge, L. W. 1961. *Gut* 2:323–37
61. Schedl, H. P., Pierce, C. E., Rider, A., Clifton, J. A., Nokes, G. 1968. *J. Clin. Invest.* 47:417–25
62. Adibi, S. A. 1969. *Gastroenterology* 56:903–13
63. Adibi, S. A. 1971. *J. Clin. Invest.* 50:2266–75
64. Matthews, D. M. 1975. *Physiol. Rev.* 55:537–608
65. Silk, D. B., Webb, J. P., Lane, A. E., Clark, M. L., Dawson, A. M. 1974. *Gut* 15:444–49
66. Keene, M. F. L., Hewer, E. E. 1929. *Lancet* 1:767–69
67. Werner, B. 1948. *Acta Paediatr. Scand.* 35:Suppl. 70, pp. 1–80
68. Hirsch-Marie, H., Loisillier, F., Touboul, J. P., Burtin, P. 1976. *Lab. Invest.* 34:623–32
69. Freeman, H., Kim, Y. 1977. *Enzymes in Gastric Cancer. Union Against Cancer Symp., Geneva.* Apr. 1977
70. Grand, R. J., Watkins, J. B., Torti, F. M. 1976. *Gastroenterology* 70: 790–810
71. Conklin, J. L. 1962. *Am. J. Anat.* 11:181–203
72. Liu, H. M., Potter, E. L. 1962. *Arch. Pathol.* 74:439–52
73. Lieberman, J. 1966. *Gastroenterology* 50:183–90
74. Zoppi, G., Andreotti, G., Pajno-Ferrara, F. 1972. *Pediatr. Res.* 6:880–86
75. Levin, R. J., Koldovsky, O., Hoskova, J., Jirsova, V., Uher, J. 1968. *Gut* 9:206–13
76. Lev, R., Siegel, H. I., Bartman, J. 1972. *Histochemie* 29:103–19
77. Henning, S. J., Kretchmer, N. 1973. *Enzyme* 15:3–23
78. Lindberg, T. 1966. *Clin. Sci.* 30:505–15
79. Jirosvá, V., Koldovský, O., Heringová, A., Hoskova, J., Jirasek, J., Uher, J. 1966. *Biol. Neonat.* 9:44–49
80. Mahboob, S., Zeman, F. J. 1976. *Nutr. Rep. Int.* 14:423–38
81. Tollolon, E. P., Goldfinger, S. E. 1974. *Surg. Clin. North Am.* 54:647–53
82. Bradley, E. L., Isaacs, J., Hersh, T., Davidson, E. D., Millikan, W. 1975. *Ann. Surg.* 182:415–29
83. Goligher, J. C., Pulvertaft, C. N., Watkinson, G. 1964. *Br. Med. J.* 1:455–60
84. Hillman, H. S. 1968. *Gut* 9:576–84
85. MacKay, C. 1970. *Am. J. Surg.* 120:324–28
86. Glazebrook, A. J., Welbourne, R. B. 1952. *Br. J. Surg.* 40:111–17
87. Annis, D., Hallenbeck, G. A. 1952. *Surgery* 31:517–27
88. Worning, H., Mulleritz, S., Hess Thaysen, E., Bang, H. O. 1967. *Scand. J. Gastroenterol.* 2:23–38
89. Lundh, G. 1958. *Acta Chir. Scand. Suppl.* 231:1–83
90. Lagerlof, H. O., Rudewald, M. B., Perman, G. 1960. *Acta. Med. Scand.* 168:269–84
91. Johnson, A. G., McDermott, S. J. 1973. *Lancet* 2:589–91
92. Go, V. L., Hoffman, A. F., Summerskill, W. H. 1970. *Gastroenterology* 58:321–28
93. MacGregor, I. L., Gueller, R., Watts, H. D., Meyer, J. H. 1976. *Gastroenterology* 70:190–96
94. MacGregor, I. L., Parent, J., Meyer, J. H. 1977. *Gastroenterology* 72:195–205

95. MacGregor, I. L., Martin, P., Meyer, J. H. 1977. *Gastroenterology* 72:206–11
96. Cowley, D. J. 1972. *Gut* 13:99–102
97. Lillibridge, C. B., Brandborg, L. L., Rubin, C. E. 1967. *Gastroenterology* 52:792–809
98. Hadorn, B., Tarlow, M., Lloyd, J. D., Wolff, O. H. 1969. *Lancet* 1:812–13
99. Tarlow, M. J., Hadorn, B., Arthurton, M. W. 1970. *Arch. Dis. Child.* 45:651–55
100. Lebenthal, E., Antonowicz, I., Shwachman, H. 1976. *Gastroenterology* 70: 508–12
101. Rutgeerts, L., Tytgat, G., Eggermont, E. 1972. *Gastroenterology* 63:381
102. Farber, S., Schwachman, H., Maddock, C. 1943. *J. Clin. Invest.* 22:827–38
103. Townes, P. L. 1965. *J. Pediatr.* 66:275–85
104. Townes, P. L., Bryson, M. F., Miller, G. 1967. *J. Pediatr.* 71:220–24
105. Hadorn, B., Hess, J., Troesch, V., Verhage, W., Götze, H., Bender, S. 1974. *Gastroenterology* 66:548–55
106. Dimagno, E. P., Go, W. L., Summerskill, W. H. 1973. *N. Engl. J. Med.* 288:813–15
107. Sarles, H., Sarles, J. C., Camatte, R., Muratore, R., Gaini, M., Guien, C., Pastor, J., Le Roy, F. 1965. *Gut* 6:545–59
108. Lundh, G. 1962. *Gastroenterology* 42:275–80
109. Wormsley, K. G. 1972. *Clin. Gastroenterol.* 1:27–51
110. Go, V. L., Hofmann, A. K., Summerskill, W. H. 1970. *J. Clin. Invest.* 49:1558–64
111. Gould, B. S., Schwachman, H. 1956. *Am. J. Dis. Child.* 91:584–87
112. Cerda, J. J., Brooks, F. P., Prockop, D. J. 1968. *Gastroenterology* 54:358–65
113. Littman, A., Hanscom, D. H. 1969. *N. Engl. J. Med.* 281:201–4
114. Saunders, J. H., Wormsley, K. G. 1975. *Gut* 16:157–62
115. Frazer, A. C. 1956. *Proc. R. Soc. Med.* 49:1009–13
116. Frazer, A. C., Fletcher, R. F., Ross, C. A., Shaw, B., Sammons, H. G., Schneider, R. 1959. *Lancet* 2:252–55
117. Frazer, A. C. 1962. *Adv. Clin. Chem.* 5:69–106
118. Messer, M., Anderson, C. M., Townley, R. R. 1961. *Clin. Chim. Acta* 6:768–75
119. Bronstein, H. D., Haeffner, L. J., Kowlessar, O. D. 1966. *Clin. Chim. Acta* 14:141
120. Cornell, H. J., Townley, R. R. 1973. *Clin. Chim. Acta* 43:113–25
121. Townley, R. R., Bhathal, P. S., Cornell, H. J., Mitchell, J. D. 1973. *Lancet* 1:1363–64
122. Lindberg, T. 1967. *Acta Paediatr. Scand.* Suppl. 177:23–24
123. Lindberg, T. 1968. *Scand. J. Gastroenterol.* 3:177–82
124. Douglas, A. P., Peters, T. J. 1970. *Gut* 11:15–17
125. Heizer, W. D., Laster, L. 1969. *J. Clin. Invest.* 48:210–28
126. Berg, N. O., Dahlquist, A., Lindberg, T., Norden, A. 1970. *Gastroenterology* 59:575–82
127. Berg, N. O., Dahlquist, A., Lindberg, T., Mansson, T., Norden, A., Rorsman, H. 1970. *Acta Dermatol.* 50:42–48
128. Kumar, V., Ghai, O. P., Chase, H. P. 1971. *Arch. Dis. Child.* 46:801–4
129. Sadikali, F. 1971. *Gut* 12:276–83
130. Berg, N. O., Dahlquist, A., Lindberg, T., Lindstrand, K., Norden, A. 1972. *Scand. J. Haematol.* 9:167–73
131. Kim, Y. S., McCarthy, D. M., Lane, W., Fong, W. 1973. *Biochim. Biophys. Acta* 321:262–73
132. Asatoor, A. M., Cheng, B., Edwards, E. D., Lant, A. F., Matthews, D. M., Milne, M. D., Navab, F., Richards, A. J. 1970. *Gut* 11:380–87
133. Hellier, M. D., Holdsworth, C. D., Perrett, D. 1972. *Clin. Sci.* 43:659–68
134. Perheentupa, J., Visakorpi, J. K. 1965. *Lancet* 2:813–16
135. Whelan, D. T., Schriver, C. R. 1968. *Pediatr. Res.* 2:525–34
136. Drummond, K., Michael, A., Ulstrom, A., Good, R. 1964. *Am. J. Med.* 37:928–48
137. Smith, A. J., Strang, L. B. 1958. *Arch. Dis. Child.* 33:109–13
138. Hooft, C., Timmermans, J., Snoeck, J., Antener, I., Oyaert, W., van den Hende, C. 1965. *Ann. Pediatr.* 205:73–84
139. Goodman, S. I., McIntyre, C. A., O'Brien, D. 1967. *J. Pediatr.* 71:246–49
140. Silk, D. B., Kumar, P. J., Perrett, D., Clark, M. L., Dawson, A. M. 1974. *Gut* 15:1–8

141. Adibi, S. A., Fogel, M. R., Agrawal, R. M. 1974. *Gastroenterology* 67:586–91
142. Hellier, M. D., Radhakrishran, A. N., Ganapathy, V., Mathan, V. I., Baker, S. J. 1976. *Gut* 17:511–16
143. Fogel, M. R., Ravitch, M. M., Adibi, S. A. 1976. *Gastroenterology* 71:729–33
144. Cook, G. C. 1973. *Br. J. Nutr.* 29:377–86
145. Cook, G. C. 1974. *Br. J. Nutr.* 32:163–67
146. Silk, D. B., Chung, Y. C., Kim, Y. S. 1976. *Gastroenterology* 70:79 (Abstr.)
147. Neale, G., Antcliff, A. C., Welbourn, R. B., Mollin, D. L., Booth, C. C. 1967. *Q. J. Med.* 36:469–94
148. Jones, E. A., Craigie, A., Tavill, A. S., Franglen, G., Rosenoer, V. M. 1968. *Gut* 9:466–69
149. Giannella, R. A., Rout, W. R., Toskes, P. P. 1974. *Gastroenterology* 67:965–74
150. Longstreth, G. F., Newcomer, A. D. 1975. *Mayo Clin. Proc.* 50:284–93

Ann. Rev. Med. 1978. 29:117–28
Copyright © 1978 by Annual Reviews Inc. All rights reserved

THE PRENATAL DIAGNOSIS OF GENETIC DISORDERS

◆7271

Charles J. Epstein
Departments of Pediatrics and Biochemistry & Biophysics, University of California, San Francisco, California 94143

Mitchell S. Golbus
Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, California 94143

INTRODUCTION

A major goal of genetic counseling is to provide reproductive options for couples who are at risk of having a child with a genetically determined disorder. Until about ten years ago, the available options were very limited and, in situations in which the risks were considered to be too high, principally included various means for avoiding conception. For the many families anxious to have children of their own, as well as for those who were counseling them, such options were clearly unsatisfactory. It was therefore with high hopes that the first and, in retrospect, quite bold attempts were made to diagnose genetic disorders prenatally. The principal objective was, and still is, to determine whether a fetus is or is not actually affected, thereby converting probabilities into certainties and permitting the family to decide whether or not to continue the pregnancy. These efforts, which were dependent upon significant advances in cell culture, cytogenetic, and obstetrical techniques, rapidly established the feasibility of prenatal diagnosis and its applicability to a wide variety of genetic problems. It has taken longer, however, to demonstrate that the act of obtaining amniotic fluid by amniocentesis is a procedure with a low enough morbidity to warrant its widespread use (1).

For the most part, prenatal diagnosis relies upon the cytogenetic or biochemical analysis of cultured amniotic fluid cells. These techniques have been described in several recent reviews (2–6) and are not further elaborated on here. In this article we present a summary of the current general experience with prenatal diagnosis based on conventional techniques, and then discuss in more detail approaches which have only recently been introduced or are in the process of development.

0066-4219/78/0401-0117$01.00

CURRENT STATUS OF PRENATAL DIAGNOSIS

Chromosomal Abnormalities

The most common indication for prenatal diagnosis is a fetus at risk of chromosomal abnormality. The relationship of the incidence of trisomy 21 to increased maternal age has been well established (7, 8), and a similar relationship has been demonstrated recently for extreme (over 55 yr) paternal age (9). The risk of having a newborn with certain other chromosomal aneuploidies (trisomy 13; trisomy 18; 47,XXY; 47,XXX; 47,XYY) also rises with increased maternal age, but to a lesser degree. There is no rigid definition of advanced maternal age, but most centers are using either 35 or 37 yr as their criterion.

Women who have previously borne a trisomic child are at a 1–2% recurrence risk of having another, irrespective of age (2, 10), and it is considered appropriate to provide prenatal diagnosis in subsequent pregnancies. Although a much smaller segment of the population is affected, the greatest risk of producing a chromosomally abnormal fetus exists for individuals who are carriers of a balanced chromosomal translocation. The most common situation is familial Down syndrome which results from the translocation of the long arm of chromosome 21 to the long arm of chromosome 14. Approximately 15% of newborns of women who carry this balanced translocation will have the Down syndrome, while the comparable risk for the offspring of men who carry this balanced translocation is 2–4% (11). Other balanced translocations or inversions are also associated with a significant risk of producing aneuploid offspring (12).

There have been 8764 pregnancies monitored at 76 centers for a cytogenetic indication (Table 1) (3). Of the 5253 fetuses studied because of advanced maternal age, 143 (2.7%) were found to be aneuploid. The number of aneuploid fetuses from pregnancies monitored because of a previous child with trisomy 21 was in the predicted range (1.2%). Although the miscellaneous indications were not detailed by all authors, most of these gestations were studied because of a history of children with the Down syndrome elsewhere in the family, a previous child with a chromosomal abnormality, or a previous child with multiple congenital anomalies. It is of interest that even with these "softer" indications, the frequency of chromosomally abnormal fetuses was as high as 2.0%.

X-Linked Disorders

A woman who is a carrier of a deleterious X-linked gene can have the sex of the fetus determined and, if she desires, abort a male fetus. At the present time, affected and unaffected males can be distinguished from one another for only three X-linked disorders, all of which are enzymatically defined: Fabry disease, the Hunter syndrome, and the Lesch-Nyhan syndrome. For all other X-linked diseases, the family must understand that there is only a 50% chance that any male fetus identified prenatally will actually have the disease in question. Fetal sex may be ascertained by demonstration of sex chromatin (Barr) body, by Y-chromosome fluorescence, and by karyotyping. Since the first two of these methods have only a 95% accuracy (13), a complete karyotype is considered mandatory. The results in Table 1 indicate that close to the expected 50% of fetuses monitored for X-linked diseases were male. One reported error in fetal sex determination occurred when sex chromatin analysis was used for sexing (14).

Biochemical Disorders

The prenatal diagnosis of a biochemical defect is generally carried out by assay of enzyme activities in cultured amniotic fluid cells. Cultured amniotic fluid cells are usually fibroblastic in their growth characteristics and express a biochemical phenotype similar to skin fibroblasts. At the present time, approximately 40 disorders have actually been diagnosed prenatally and another 35 are capable of being diagnosed [see (3) or (4) for compilation]. The combined data in Table 1 indicate that 522 pregnancies have been monitored for an inborn error of metabolism; of these, 126 (24.1%) fetuses were affected. Tay-Sachs disease, Pompe disease, and the mucopolysaccharidoses have been the most common indications for prenatal biochemical studies thus far.

From the public health point of view, our current handling of the autosomal recessive conditions is unsatisfactory, since even widespread application of prenatal diagnosis *after* the birth of an abnormal child will provide only a limited reduction in disease incidence. In an attempt to obviate this problem, programs of mass screening for heterozygote detection are being developed. The criteria for such screening programs include the following: a defined population at risk; an accurate, reliable, and inexpensive method of carrier detection; and the ability to prenatally diagnose the entity in question. Tay-Sachs was the first disorder to meet these criteria, and a worldwide network of Tay-Sachs screening

Table 1 Cumulative experience of prenatal detection centers

Indication	Pregnancies studied	"Affected" fetuses (No.)	"Abnormal" fetuses (%)
Chromosomal defects			
translocation carriers	293	29	9.9
maternal age[a] ≥35 years	3,012	79	2.6
≥40 years	957	42	4.4
35–39 years	1,284	22	1.7
previous trisomy 21	1,924	23	1.2
miscellaneous	1,294	26	2.0
X-linked diseases	433	200[b]	46.2 (males)
Biochemical defects	522	126	24.1
Neural tube defects	1,600	98[c]	5.5[d]
Totals	11,319	645	5.7

[a] The pregnancies are divided into three separate groups because of the methods of reporting employed by different authors.

[b] Includes one trisomy 21 fetus.

[c] Includes 10 fetuses with NTD found at routine screening of fluid drawn for another genetic indication.

[d] When indication was for a neural tube defect.

programs has been established (15). Over 150,000 individuals have been tested, and a heterozygote frequency of 1:27 has been found in the Ashkenazi Jewish population (M. M. Kaback, personal communication). Approximately 400 at-risk pregnancies (in which both parents were heterozygotes) have already been monitored and 300 unaffected offspring born. With the recent prenatal diagnosis of a number of hemoglobinopathies (see below), the feasibility of approaching this class of disorders with similar mass screening efforts appears to be greater than had originally been thought.

Neural Tube Defects

The discovery by Brock & Sutcliffe (16) that increased quantities of α-fetoprotein are present in amniotic fluid and maternal serum in pregnancies in which the fetus had a neural tube defect (NTD) has led to the prenatal diagnosis of a number of such fetuses. α-Fetoprotein is synthesized by the fetal liver and yolk sac and normally reaches its highest concentrations in fetal serum and amniotic fluid during the 12th and 14th gestational weeks. The increase in α-fetoprotein in NTD is presumably due to transudation from exposed fetal capillaries. Individuals with NTD are at a 2–5% risk, depending on genetic background and geographic location, of having siblings or children with a similar defect; these have been the families referred for prenatal diagnosis. Table 1 indicates 1600 such at-risk pregnancies that have been monitored, and 88 affected fetus have been found. However, increased amniotic fluid α-fetoprotein level is not in itself specific for NTD and has been found in conjunction with a number of other congenital defects such as sacrococcygeal teratoma, nephrosis, omphalocele, and esophageal atresia.

Of newborns with NTD, 90% are the first affected individual in a family. Therefore, even if amniocentesis were employed in all families with a previously affected child, the disease incidence would be lowered by only 10%. This has stimulated an interest in assaying α-fetoprotein levels in maternal serum in all pregnancies. Approximately 80% of fetuses with anencephaly and 60% with open spina bifida are associated with maternal serum α-fetoprotein levels above the highest value for normal pregnancies (17). The resultant overall detection rate of 68% suggests that the disease incidence could be more than halved by widespread application of maternal serum screening. However, if the 98th percentile of the normal range of values were used as an indication for further evaluation, including a confirmatory amniocentesis and determination of amniotic fluid α-fetoprotein concentration, this would generate 60,000 additional amniocenteses annually in the United States, an amount far exceeding our current capabilities. Nevertheless, maternal serum screening programs are being established in the United Kingdom where the incidence of NTD seems high enough to warrant this approach.

FETAL VISUALIZATION

Most multifactorially inherited congenital malformations and many Mendelian disorders occur without demonstrable chromosomal or biochemical abnormalities. The prenatal diagnosis of such entities has been attempted by either indirect or direct fetal visualization. Indirect methods include sonography and radiography. With the introduction of gray-scale and real-time sonography, a number of congenital anomalies have already been prenatally diagnosed. These include anencephaly, myelomeningocele, omphalocele, and polycystic kidneys. In some cases, serial measurements of fetal ventricular size to evaluate hydrocephalus and microcephaly appear to be of value in diagnosing these abnormalities in the first half of pregnancy. Contrast radiography with a radiopaque water-soluble dye such as Renografin has been used to demonstrate fetal swallowing and thereby the presence of esophageal or duodenal atresia, and to outline the fetal silhouette to detect an abnormal mass or short-limbed dwarfism.

Direct radiography has been used to demonstrate fetal radii and, therefore, to determine that a fetus was not affected by the autosomal recessively inherited syndrome of thrombocytopenia and absent radii (18). It has also been used to make the prenatal diagnosis of Saldino-Noonan dwarfism (A. Beaudet, personal communication). However, x rays of an achondroplastic fetus and of a fetus with infantile osteopetrosis, both at 20 wk gestation, did not result in the diagnosis of existing disease (19; M. S. Golbus, B. D. Hall, & R. A. Filly, unpublished data). The use of direct radiography for prenatal diagnosis thus appears to be quite limited, and the use of both direct and contrast radiography will probably be superseded by the development of better sonographic techniques.

Direct fetal visualization is performed using a small-bore fiberoptic endoscope. Practicality dictated that a fetoscopic method be developed which can be performed as an outpatient procedure under local anesthesia. This has been made possible by the production of a 1.7-mm diameter endoscope (Needlescope-Dyonics Inc., Woburn, Mass.) containing a solid lens surrounded by fiber bundles transmitting the light. The introducing cannula has an outer diameter of 2.2 mm and accepts a sharp trocar for transabdominal insertion into the amniotic sac. Initial use of the fetoscope in 283 volunteers undergoing elective second trimester abortions with saline or prostaglandin $F_{2\alpha}$ resulted in successful entry into the amniotic sac in 271 instances (20, 21, and our unpublished data). Although isolated parts of the fetal anatomy are sporadically identified, visualization of the entire fetus is rarely accomplished. Nevertheless, four pregnancies at risk of either the Ellis-van Creveld syndrome (an autosomal recessive condition which includes chondrodysplasia, ectodermal dysplasia, congenital heart defects, and polydactyly) or of an autosomal dominant split-hand syndrome have been examined fetoscopically. On the basis of visualization of polydactyly, one fetus was diagnosed as having the Ellis-van Creveld syndrome, and this was verified when the pregnancy was terminated (20). However, the major use of fetoscopy has been for directly visualized sampling of fetal blood from the placenta for the prenatal diagnosis of hemoglobinopathies. Of 40 pregnancies which were subjected to such endoscopic sampling with the intention of being allowed to proceed to term, 5 terminated in spontaneous abortions (J. C. Hobbins, personal communication).

The fetoscope could conceivably also be used for direct sampling of other tissues such as placenta and skin, or even of internal fetal organs. Although it might be possible to obtain enough fibroblasts for karyotype analysis or biochemical studies somewhat more rapidly from skin explants, the time saved does not seem to be worth the risk to the fetus. Muscle biopsies, however, might be of considerable aid in the prenatal diagnosis of myopathies.

NEWLY DEVELOPED AND DEVELOPING BIOCHEMICAL TECHNIQUES

Micromethods for Enzyme Assay

A limiting factor in the diagnosis of hereditary enzyme defects by conventional techniques is the length of time required to grow sufficient cells for enzyme analysis. Depending on the condition being looked for, anywhere from four to eight weeks may be necessary (22). Although the relaxation of regulations regarding abortion have made this long period between amniocentesis and final answer less of a problem, at least from the legal point of view, two months is an inordinately long time for a family to wait before learning the status of their fetus. To circumvent this problem, more sensitive techniques for enzyme analysis are being developed, and some have already been put into practice. These methods, developed by Galjaard and collaborators (23) and by Hösli (24), are based mainly on the measurement of enzyme activity in small numbers of cells (of the order of tens to thousands) using quantitative microspectrophotometry or microfluorometry. Prenatal diagnoses which used these techniques have been attempted in fetuses at risk of Pompe disease (glycogenosis II), Fabry disease, and GM_1-gangliosidosis; a fetus affected with the first of these has been detected (23, 25). In each instance it was possible to obtain definitive results within two weeks after amniocentesis. Assays of virtually all of the clinically important lysosomal enzymes, which make use of fluorescent 4-methylumbelliferyl substrates, are now feasible. In addition, enzymes such as hypoxanthine guanine phosphoribosyltransferase, the assay of which is based on the use of radioactive substrates (26), can also be determined with this micromethodology.

Utilization of Materials Other Than Cultured Cells

Another approach to avoiding the problems inherent in the biochemical analysis of cultured amniotic fluid cells is to employ techniques which can be directly applied to uncultured cells, amniotic fluid, fetal red cells, fetal serum, or to maternal blood or urine. One application of this approach has already been described with reference to the analysis of amniotic fluid and maternal serum for α-fetoprotein in the prenatal diagnosis of NTD and other structural defects.

Although originally attempted in the period shortly after the introduction of prenatal diagnosis for inherited metabolic disorders, the use of uncultured amniotic fluid cells and of cell-free amniotic fluid has generally not been favored. With uncultured amniotic fluid cells, the variable nature of the cell population obtained at amniocentesis has made it difficult to standardize the assays accurately enough to warrant their use, although reliable results have apparently been obtained for Tay-Sachs disease (27). Similarly, the examination of amniotic fluid, generally more for metabolic products than for enzymes themselves, has often proved unreliable. Despite these early difficulties, it appears that such material can be used if properly controlled. Thus, it has recently been possible to make the in utero diagnosis of vitamin B_{12}-responsive methylmalonic acidemia by gas chromatographic analysis of methylmalonic acid in the amniotic fluid (28). The observed increase in amniotic fluid methylmalonic acid was also paralleled by a marked increase in the excretion of this substance by the pregnant mother. Measurements of enzyme activities in cell-free amniotic fluid have also been reliable in the diagnosis of the Hunter syndrome (mucopolysaccharidosis II) (29), Sandhoff disease (30), Tay-Sachs disease (M. M. Kaback, personal communication), and argininosuccinic aciduria (31). While investigators may still wish to confirm the amniotic fluid results by analysis of cultured cells, the rapid initial answer is still quite useful for the family and physicians.

"Reverse" (3,3′,5′) triiodothyronine has recently been identified as a major product of fetal thyroid activity and is found in amniotic fluid in high concentrations during midpregnancy. It has been suggested that assay of this material will provide a useful assessment of fetal thyroid funtion (32) and, inferentially, that it may make possible the diagnosis of various hereditary and congenital hypothyroid states. One fetus with hypothyroidism has already been diagnosed and treated prenatally (33).

Major efforts have recently been directed toward obtaining fetal erythrocytes and serum for the analysis of hemoglobin and other constituents of blood. Experience with fetal erythrocytes obtained by aspiration from the placenta with fetoscopic or sonographic monitoring has sufficiently demonstrated the feasibility of diagnosing several disorders of hemoglobin prenatally. Utilizing sensitive radioisotopic methods for the quantitation of hemoglobin chain synthesis, it has been possible to diagnose homozygous β-thalassemia in several at-risk fetuses (34, 35). Similarly, the diagnosis of sickle cell anemia has been made on the basis of the synthesis of β^S but not of β^A chains in fetal erythrocytes. Because of the very small quantities of erythrocytes available for study, impeccable analytical technique is necessary if reliable results are to be obtained. In the initial series of cases, a significant fetal mortality was associated with the attempts to obtain fetal blood. However, as with many new methods, the earliest trials were associated with many unforeseen difficulties, which have been prevented by changes in technique, and the safety of the procedure has improved with experience.

It is likely that fetal red cells will be useful for more than the analysis of hemoglobin. Several hereditary defects of erythrocyte enzymes or structural components result in hemolytic anemias which can, in theory, be diagnosed prenatally if fetal erythrocytes can be shown to be similar to adult cells with regard to these constituents. Similarly, fetal erythrocytes may contain enzymes which are not present in amniotic fluid or amniotic fluid cells but are representative of enzymes present in other tissues. A case in point is arginase, an enzyme deficient in one type of hyperargininemia associated with progressive neurologic deterioration. Although arginase deficiency in the liver is presumed to be the cause of the disorder, the enzyme can be assayed in fetal erythrocytes, thereby making possible the prenatal diagnosis of this disorder (36).

If pure samples of fetal blood can be obtained it will be possible to obtain other fetal blood elements, as well as plasma or serum, at the same time. Availability of fetal blood would facilitate the prenatal diagnosis of conditions such as the Chediak-Higashi syndrome and chronic granulomatous disease which affect polymorphonuclear leukocyte function, of the several disorders of immune function that result in aberrations of lymphocyte development and function, and possibly even of hereditary platelet defects that occur either alone or in combination with other disorders (such as hereditary nephritis) (37). Availability of fetal plasma or serum would make possible the prenatal diagnosis of a wide variety of disorders associated with defects of plasma or serum proteins. There is already evidence that fetuses synthesize a wide variety of genetically important proteins early in development. These include the several immunoglobulins, C′1 esterase inhibitor (defective or missing in hereditary angioneurotic edema), ceruloplasmin (decreased in Wilson's disease), α_1-antitrypsin (defective or absent in certain hereditary forms of chronic emphysema and hepatic cirrhosis), fibrinogen (abnormal in various dysfibrinogenemias and absent in afibrinogenemia), and complement (38, 39). Of particular note, since prenatal diagnosis for it is frequently requested, is factor VIII, the clotting constituent which is defective in hemophilia A. This factor can be detected both antigenically and by activity in fetuses as young as 14–16 wk gestational age (40).

Another condition for which prenatal diagnosis is relatively frequently sought is Duchenne muscular dystrophy. Like hemophilia A, it is X-linked, and the current approach is to determine the sex of the fetus. However, since Duchenne muscular dystrophy is known to result postnatally in very high activities of serum creatine kinase, presumably due to leakage of the enzyme from diseased muscle, efforts are now being made by ourselves and others (41) to determine whether assay of this enzyme prenatally will be useful for prenatal diagnosis.

Diagnosis of Conditions Not Detectable in Readily Available Materials or by Visualization Techniques

Despite the large number of enzyme defects and other hereditary and congenital abnormalities that can be identified directly or indirectly in amniotic fluid cells, amniotic fluid, and fetal blood, or that can be visualized by one of several physical techniques, there are still a great number of genetic disorders which cannot be so diagnosed. These approaches are of no value either because the biochemical defects are not detectable in the available materials, because the underlying defects have not been identified, or because there are no easily visible structural defects. Efforts are therefore being made to develop other approaches to prenatal diagnosis, and prototypes for two of these new approaches now exist.

A fundamental, although still unproven, tenet of developmental genetics is that the genomes of all cells from an individual are identical even though the cells from which they are derived have quite divergent properties. Therefore, if a hereditary defect can be identified at the level of the gene(s) or DNA sequences involved, DNA from any tissue should be suitable for analysis. The validity of this approach has now been established in the prenatal diagnosis of α-thalassemia (42). This disorder is produced by an absence of hemoglobin α-chain synthesis which, in turn, is the result of a deletion of all four α-chain genes from the genome. Although α-chains are not normally made by amniotic fluid cells, the genes which govern their synthesis are present. Therefore, it has been possible, using molecular hybridization techniques, to detect α-thalassemia-1 (two genes deleted), hemoglobin H disease (three genes deleted), and α-thalassemia (four genes deleted) by analysis of DNA derived from cultured amniotic fluid cells (42). This approach should be applicable to any genetic disorder in which the defect results from the deletion of a gene for which the specific messenger RNA can be obtained or synthesized. Although these requirements have not been met for any nonhemoglobin related conditions, the recent explosion in molecular genetic research suggests that they ultimately will be.

Another way, at least in principle, to make use of the presence of unexpressed genes in amniotic fluid cells is to develop methods for inducing the expression of these genes. A precedent for this approach now exists in somatic cell hybridization experiments in which nuclei from fibroblasts or leukocytes were observed to inititate the synthesis of albumin, a liver cell protein, after fusion with hepatoma cells (43, 44). Successful development of methods for the activiation of unexpressed genes would be particularly useful for the prenatal diagnosis of conditions such as urea cycle defects, certain mucopolysaccharidoses, muscle abnormalities, phenylketonuria, and the like, which cannot presently be approached.

An entirely different approach to the diagnosis of conditions for which available materials or molecular techniques are not suitable is the utilization of linked genetic markers. With increasing frequency, genes for readily detectable biochemical or immunologic markers that are closely linked to genes for inherited disorders are being discovered (45). In favorable situations, when the carrier of a deleterious gene is also heterozygous for such a closely linked marker and when the mate is of an appropriate genotype, it may be possible to follow the transmission of the gene to a fetus indirectly by actually following the transmission of the marker. This approach has already been applied to the prenatal diagnosis of myotonic dystrophy (46, 47). The gene for this disorder is closely linked to the secretor locus that controls the secretion of ABH blood group substance. Other potentially useful linkages for prenatal diagnoses include glucose-6-phosphate dehydrogenase (G-6-PD) and hemophilia A (particularly in blacks, who have a high frequency of G-6-PD variants), HLA and the various complement loci, and HLA and dominantly inherited spinocerebellar ataxia. The last of these linkages (48), which is reasonably tight, may make possible the prenatal diagnosis of a condition that does not ordinarily become manifest until adult life.

CONCLUSION

Prenatal diagnosis has proven highly effective in assessing the status of fetuses at risk of all cytogenetic and of several biochemical and structural disorders with genetic etiologies. The original methodology, based on the cytogenetic and biochemical analysis of cultured amniotic fluid cells, has now been complemented by a wide variety of new techniques. These include several methods for fetal visualization (sonography, fetoscopy, x ray), sampling of fetal blood, and analysis of very small quantities of material. Additional approaches, based on these and other technologies, are likely to permit the prenatal diagnosis of an ever increasing number of genetic disorders. At the same time, the number of pregnancies monitored will increase as greater resources become available and as screening programs identify couples at risk of having genetically abnormal children prior to the birth of an affected child. Prenatal diagnosis is already a powerful means of preventing the birth of individuals with significant genetic defects, thereby sparing both individuals and society from the burdens that such disorders produce. In the future, it is likely to be even more effective.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Institutes of Health (GM-19527), the Maternal and Child Health Service (Project No. 445), and the National Foundation—March of Dimes. C. J. E. is an Investigator of the Howard Hughes Medical Institute.

Literature Cited

1. NICHD National Registry for Amniocentesis Study Group. 1976. Midtrimester amniocentesis for prenatal diagnosis. Safety and accuracy. *J. Am. Med. Assoc.* 236:1471–77
2. Milunsky, A. 1973. *The Prenatal Diagnosis of Hereditary Diseases,* Springfield, Ill: Thomas. 253 pp.
3. Golbus, M. S. 1977. The prenatal diagnosis of genetic defects. In *Advances in Obstetrics and Gynecology,* ed. R. M. Caplan, W. J. Sweeney, pp. 88–104 Baltimore: Williams & Wilkins. 288 pp.
4. Epstein, C. J., Golbus, M. S. 1977. Prenatal diagnosis of genetic diseases. *Am. Sci.* In press
5. Epstein, C. J. 1976. Prenatal diagnosis of genetic diseases. In *Clinical Chemistry, ACS Symp. Ser. 36,* ed. D. T. Forman, R. W. Mattoon, pp. 69–94. Washington, DC: Am. Chem. Soc. 293 pp.
6. Boué, A. ed. 1976. *Prenatal Diagnosis,* Vol. 61. Paris: INSERM. 330 pp.
7. Jenkins, R. 1933. Etiology of mongolism. *Am. J. Dis. Child.* 45:506–19
8. Penrose, L. S. 1933. The relative effects of paternal and maternal age in mongolism. *J. Genet.* 27:219–24
9. Stene, J., Fischer, G., Stene, E., Mikkelsen, M., Peterson, E. 1977. Paternal age effect in Down's syndrome. *Ann. Hum. Genet.* 40:299–306
10. Mikkelson, M., Stene, J. 1970. Genetic counselling in Down's syndrome. *Hum. Hered.* 20:457–64
11. Polani, P. E. 1974. Prenatal cytological recognition of sex-linked and chromosomal abnormalities. *J. Obstet. Gynaecol. Br. Commonw.* 78:1024–30
12. Boué, J., Boué, A. 1976. Prenatal diagnosis in carriers of balanced chromosomal structural anomalies. In *Proc. 5th Int. Congr. Hum. Genet., Int. Congr. Ser. No. 397,* ed. S. Armendares, R. Lisker, p. 112 (Abstr.). Amsterdam & Princeton, NJ: Excerpta Med.
13. Nadler, H. L. 1971. Indication for amniocentesis in the early prenatal detection of genetic disorders. *Birth Defects, Original Article Series 7(5):5–9*
14. Fuchs, F. 1966. Genetic information from amniotic fluid constituents. *Clin. Obstet. Gynecol.* 9:565–73
15. Kaback, M. M., Zeiger, R. S., Reynolds, L. W., Sonneborn, M.

1974. Approaches to the control and prevention of Tay-Sachs disease. *Prog. Med. Genet.* 10:103–34

16. Brock, D. J. H., Sutcliffe, R. G. 1972. Alpha-fetoprotein in the antenatal diagnosis of anencephaly and spina bifida. *Lancet* 2:197–99
17. Laurence, K. M. 1974. Clinical and ethical considerations on alpha-fetoprotein estimation for early prenatal diagnosis of neural tube malformations. *Dev. Med. Child. Neurol.*, Suppl. 32:117–21
18. Omenn, G. S., Figley, M. M., Graham, C. B., Heinrichs, W. L. 1973. Prospects for radiographic intrauterine diagnosis—The syndrome of thrombocytopenia with absent radii. *N. Engl. J. Med.* 288:777–78
19. Golbus, M. S., Hall, B. D. 1974. Failure to diagnose achondroplasia in utero. *Lancet* 1:629
20. Benzie, R. J., Doran, T. A. 1975. The "fetoscope"—A new clinical tool for prenatal diagnosis. *Am. J. Obstet. Gynecol.* 121:460–64
21. Hobbins, J. C., Mahoney, M. J. 1976. Fetoscopy and fetal blood sampling: the present state of the method. *Clin. Obstet. Gynecol.* 19:341–52
22. Galjaard, H., Fernandes, J., Jahodova, M., Koster, J. F., Niermeijer, M. F. 1972. Prenatal diagnosis of genetic disease. *Bull. Eur. Soc. Hum. Genet.* Nov., pp. 79–91
23. Galjaard, H., Hoogeveen, A., Keijzer, W., DeWit-Verbeek, E., Vlek-Noot, C. 1974. The use of quantitative cytochemical analyses in rapid prenatal detection and somatic cell genetic studies of metabolic diseases. *Histochem. J.* 6:491–509
24. Hösli, P. 1974. Microtechniques for rapid prenatal diagnosis in early pregnancy. In *Birth Defects*, ed. A. G. Motulsky, W. Lenz, pp. 226–33. Amsterdam: Excerpta Med. 373 pp.
25. Kleijer, W. J., Van der Veer, E., Niermeijer, M. F. 1976. Rapid prenatal diagnosis of GM$_1$-gangliosidosis using microchemical methods. *Hum. Genet.* 33:299–305
26. Singh, S., Willers, I., Goedde, H. W. 1976. A rapid micromethod for prenatal diagnosis of Lesch Nyhan syndrome. *Clin. Genet.* 10:12–15
27. O'Brien, J. S., Okada, S., Fillerup, D. L., Veath, M. L., Adornato, B., Breuner, P. H., Leroy, J. G. 1971. Tay-Sachs disease: prenatal diagnosis. *Science* 172:61–64
28. Nakamura, E., Rosenberg, L. E., Tanaka, K. 1976. Microdetermination of methylmalonic acid and other short chain dicarboxylic acids by gas chromatography: use in prenatal diagnosis of methylamalonic acidemia and in studies of isovaleric acidemia. *Clin. Chim. Acta* 68:127–40
29. Liebaers, I., DiNatale, P., Neufeld, E. F. 1977. Iduronate sulfatase in amniotic fluid: An aid in the prenatal diagnosis of the Hunter syndrome. *J. Pediatr.* 90:423–25
30. Desnick, R. J., Krivit, W., Sharp, H. L. 1973. *In utero* diagnosis of Sandhoff's disease. *Biochem. Biophys. Res. Commun.* 51:20–26
31. Fleisher, L. D., Rassin, D. K., Rogers, P., Desnick, R. J., Gaull, G. E. 1977. Argininosuccinic aciduria: prenatal diagnosis and studies of an affected fetus. *Pediatr. Res.* 11:455 (Abstr.)
32. Chopra, I. J., Crandall, B. F. 1975. Thyroid hormones and thyrotropin in amniotic fluid. *N. Engl. J. Med.* 293:740–43
33. Van Herle, A. J., Young, R. T., Fisher, D. A., Uller, R. P., Brinkman, C. R. III. 1975. Intrauterine treatment of a hypothyroid fetus. *J. Clin. Endocrinol. Metab.* 40:474–77
34. Kan, Y. W., Golbus, M. S., Trecartin, R. F., Filly, R. A., Valenti, C., Furbetta, M., Cao, A. 1977. Prenatal diagnosis of β-thalassemia and sickle-cell anaemia. Experience with 24 cases. *Lancet* 1:269–71
35. Alter, B. P., Modell, C. B., Fairweather, D., Hobbins, J. C., Mahoney, M. J., Frigoletto, F. D., Sherman, A. S., Nathan, D. G. 1976. Prenatal diagnosis of hemoglobinopathies. A review of 15 cases. *N. Engl. J. Med.* 295:1437–43
36. Spector, E. B., Cederbaum, S. D., Bernard, B. 1977. Properties of human adult and fetal red blood cell arginase: a possible diagnostic test for arginase deficiency. *Pediatr. Res.* 11:464 (Abstr.)
37. Epstein, C. J., Sahud, M. D., Piel, C. F., Goodman, J. C., Bernfield, M. R., Kushner, J. H., Ablin,

A. R. 1972. Hereditary macrothrombocytopathia, nephritis and deafness. *Am. J. Med.* 52:299–310

38. Gitlin, D., Biasucci, A. 1969. Development of γG, γA, γM, β1C/β1A, C′1 esterase inhibitor, ceruloplasmin, transferrin, hemopexin, haptoglobin, fibrinogen, plasminogen, α-antitrypsin, orosomucoid, β-lipoprotein, α_2-macroglobulin, and pre-albumin in the human conceptus. *J. Clin. Invest.* 48:1433–46
39. Kohler, P. F. 1973. Maturation of the human complement system. I. Onset time and sites of fetal $C1_q$, C4, C3 and C5 synthesis. *J. Clin. Invest.* 53:671–77
40. Holmberg, L., Henriksson, P., Ekelund, H., Åstedt, B. 1974. Coagulation in the human fetus. Comparison with term newborn infants. *J. Pediatr.* 85:860–64
41. Haseltine, F. P., Mahoney, M. J., Hobbins, J. C., Caskey, C. T. 1977. Possibility of prenatal diagnosis of Duchenne muscular dystrophy (DMD). *Pediatr. Res.* 11:457 (Abstr.)
42. Kan, Y. W., Golbus, M. S., Dozy, A. M. 1976. Prenatal diagnosis of α-thalassemia. Clinical application of molecular hydridization. *N. Engl. J. Med.* 295:1165–67
43. Peterson, J. W., Weiss, M. C. 1972. Expression of differentiated functions in hepatoma cell hybrids: Induction of mouse albumin production in rat hepatoma-mouse fibroblast hybrids. *Proc. Natl. Acad. Sci. USA* 69:571–75
44. Darlington, G. J., Bernhard, H. P., Ruddle, F. H. 1974. Human serum albumin phenotype activation in mouse hepatoma-human leukocyte cell hybrids. *Science* 185:859–62
45. McKusick, V. A., Ruddle, F. H. 1977. The status of the gene map of the human chromosomes. *Science* 196:390–405
46. Schrott, H. G., Karp, L., Omenn, G. S. 1973. Prenatal prediction in myotonic dystrophy: Guidelines for genetic counseling. *Clin. Genet.* 4:38–45
47. Insley, J., Bird, G. W. G., Harper, P. S., Pearse, G. W. 1976. Prenatal prediction of myotonic dystrophy. *Lancet* 1:806
48. Jackson, J. F., Currier, R. D., Terasaki, P. I., Morton, N. E. 1977. Spinocerebellar ataxia and HLA linkage. Risk prediction by HLA typing. *N. Engl. J. Med.* 296:1138–41

Ann. Rev. Med. 1978. 29:129–36
Copyright © 1978 by Annual Reviews Inc. All rights reserved

E. COLI MENINGITIS: K_1 ANTIGEN AND VIRULENCE

◆7272

Catherine M. Wilfert, M.D.
Associate Professor Pediatrics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710

INTRODUCTION

Meningitis due to *Escherichia coli* occurs primarily in newborn infants and as a nosocomial infection in hospitalized adults. The mortality of these infections has been reported to range from 40–80% in newborns, with the majority of surviving infants manifesting subsequent neurological or developmental effects of the infection. The poor prognosis led to a prospective controlled evaluation of intrathecal antibiotic therapy in newborn infants initiated by McCracken (1). As a result of information accumulated during this study, the relationship of capsular (K_1) antigen of *E. coli* to the invasiveness of this organism in neonates was substantiated. The present chapter reviews K_1 antigen and invasiveness of *E. coli,* host factors contributing to the pathogenesis of *E. coli* meningitis in newborns, the current epidemiology of this organism, and antimicrobial therapy.

HOST FACTORS

Premature infants are at four times greater risk from meningitis than full term infants (2), and meningitis is more common in the first month of life for full term infants than in any succeeding months (3). Retrospective analysis of infants with gram-negative infection seen at Duke University Medical Center is in accord with these data (4).

It has long been recognized that newborn infants do not receive transplacental IgA or IgM from their mothers. The IgM deficiency results in normal newborn infants having no humoral antibodies to the somatic (O) antigens of Enterobacteriaceae despite their demonstrated ability to synthesize immunoglobulins early in intrauterine life. These IgM antibodies are more efficient than IgG antibodies in their opsonization and killing effects against gram-negative bacteria (5). Infants are also deficient in several components of the complement

0066-4219/78/0401-0129$01.00

system (C_1, C_3, C_4, C_5) (6, 7) and, consequently, have a decreased activation of complement by the classical complement pathway. A deficiency of factor B (glycine-rich β-glycoprotein or C_3 proactivator) has been demonstrated in 15% of cord sera obtained from normal infants (8); factor B is necessary for the activation of the alternate complement pathway. The neonate with an absence of specific antibodies to invading organisms, deficient complement, plus a deficiency of factor B is less able to opsonize and lyse bacteria.

Attempts to study apparent leucocyte dysfunction have revealed that newborn serum has a decreased ability to generate chemotactic activity (9) and that newborn leucocytes respond less well than adult cells to chemotactic factors (9, 10). Neonatal polymorphonuclear cells phagocytize less well than adult cells in lower concentrations of serum (11). Although cells from newborns with demonstrable clinical abnormalities may phagocytize less well (12), there does not seem to be a defect in the ability of polymorphonuclear cells to kill ingested bacteria.

Finally, investigations of maternal breast milk have indicated that a number of factors that afford some protection against the onslaught of environmental flora may be passively transferred to an infant. The factors present in breast milk include antibodies primarily of the IgA class (13, 14) which specifically includes an antibody to organism such as *E. coli* (15). Mononuclear white cells are present in numbers approximating 2300/mm^3 (16–18). Others factors, nonspecific in their protective action, include a growth factor for lactobacillus bifidus (19, 20), lysozyme (21, 22), and lactoferrin (23, 24). Lactoferrin is present in concentrations of 100 mg/100 ml, is capable of inhibiting bacterial growth, and in concert with specific antibodies demonstrates an enhanced bacteriostatic effect on *E. coli* in vitro (25). Thus, the failure to breast feed may deprive an infant of this passively acquired maternal protection.

E. COLI K_1 ANTIGEN AND VIRULENCE

The genus *E. coli* is antigenically complex. There are at least 155 somatic (O), 100 capsular (K), and 50 flagellar (H) antigens. The O antigens occur in all Enterobacteriaeceae and are complex lipopolysaccharides. The thermostable polysaccharide moiety is responsible for the O specificity. The original term of K (Kapsel, in German) antigen was accepted some 30 years ago and designated either an envelope or a surface antigenic structure (26).

The *E. coli* K polysaccharides appear to be linear co- or homopolymers composed of one or several monosaccharides joined by glycosidic and/or phosphodiester bonds. They have a negative charge and a neutral pH because of these bonds. The relationship between K polysaccharides and the invasive abilities of *E. coli* was first shown by Smith (27). In 1970 it was reported that an *E. coli* isolated from the spinal fluid of a newborn was agglutinated by antiserum raised to group B meningococci (28). This *E. coli* was strain 07: K_1 (L): NM which shared a K polysaccharide antigen with group B meningococci (29). The

K_1 polysaccharide is a polymer of N-acetylneuraminic acid (30) and is immunologically identical in Group B meningococci and *E. coli*.

Experimentally, K_1 antigen–containing strains of *E. coli* have been shown to have enhanced virulence after intraperitoneal or intracerebral injection into mice (31). Passive protection was achieved in the mouse model by administration of anticapsular antibodies. In addition, the K_1 organisms isolated from cerebrospinal fluid (CSF) of newborns demonstrated this enhanced virulence (32). This recognition that a single capsular serotype of *E. coli* is the predominant gram-negative pathogen of neonatal meningitis has stimulated research to find a means to prevent such infection.

K_1 *E. COLI* AND NEONATAL MENINGITIS

K_1 *E. coli* were responsible for 70% (82 of 117) of cases of neonatal meningitis (1). A similar predominance of organisms with K_1 polysaccharide antigen (77% or 126 of 163) was reported in the collaborative study (which included infants) as well as in surveys of other institutions and countries (33).

The significance of this prevalence of K_1 antigen is emphasized by reviewing the surveys of *E. coli* isolates obtained from a variety of patients (Table 1). The quantity of K_1 antigen is similar whether the strain is isolated from the CSF, the blood of a neonate, or the stools of a healthy individual. Four somatic antigens (O18-ac, O7, O1, and O16) were associated with two thirds of the CSF K_1 strains studied (34). The predominance of relatively few serotypes is interesting although there is no clear relationship between virulence and either O or H antigens.

Table 1 K_1 polysaccharide associated with neonatal *E. coli*–caused meningitis[a]

Source	Total Cases (No.)	Cases with K_1 (No.)	Cases with K_1 (%)
CSF (neonatal meningitis)	163	126	77
Blood (septicemia)			
Neonatal (no meningitis)	103	37	36
Adult (no meningitis)	301	43	11
Urine (urinary tract infection)			
Adult	110	13	11
Children	75	14	19
Prenant women	209	23	11
Stool			
Normal adult	118	17	13
Normal 4-day-old infant	100	12	12
Enteropathogenic strains	30	0	0

[a] Modified from Sarff et al (33) and from Robbins et al (32).

Epidemiological surveys (33) suggest the following. Vertical transmission of the organism from mother to infant is the principal means of *E. coli* K_1 infection in healthy neonates. Two thirds of newborns delivered to *E. coli* K_1-positive mothers had K_1 strains in their stools. The K_1 strains tended to be the predominant organisms in infants who were culture-positive. Vertical transmission was substantiated by additional serotyping which showed that 88% of the *E. coli* strains cultivated from 51 mother-infant pairs had identical O and H serotypes. Only 11% of infants born to K_1-negative mothers demonstrated K_1 colonization and apparently had acquired their organisms later in the neonatal period. Their strains could be matched to *E. coli* K_1-bearing organisms obtained from rectal swabs of nursery staff. Of 71 infants in an intensive care nursery, 39 acquired the organism after their first week of life. Half of the nursery staff attending these infants had identical serotypes of *E. coli* in their stools. These studies imply that cross infection from personnel may be maintaining several *E. coli* serotypes within a given nursery.

The morbidity and mortality in infants sustaining meningitis with K_1-bearing *E. coli* is related to the amount of K_1 antigen detected and the length of time it is present in the serum and/or CSF (34, 35). The quantitation of antigen may only be a reflection of the number of organisms present.

There are parallels among K_1-bearing *E. coli* and other organisms causing purulent meningitis in childhood. The pathogens, *Hemophilus influenza* type b, *Neisseria meningitidis*, *Streptococcus pneumoniae*, and K_1-bearing *E. coli* are all encapsulated with polysaccharide. Anticapsular antibodies are protective against these infections in animal models which raises the possibility that antibody presence in the mother affords passive protection to the infant.

Diagnosis of E. coli Meningitis in the Neonate

The commonest signs and symptoms of neonatal bacterial meningitis at presentation are nonspecific. Lethargy, abnormal temperature, poor feeding, and irritability are the most frequent indicators of illness. Neurological signs and symptoms usually appear after the onset of the disease (2).

Laboratory diagnosis depends primarily upon the examination of the CSF. White blood cells may be present in normal newborn's CSF (36) and protein values are frequently elevated (36, 37). Approximately 4% of the infants with gram-negative meningitis studied by the collaborative study had normal CSF cell counts on the initial lumbar puncture (1). The protein value was normal in about 20% of these babies and the CSF glucose and serum glucose values were normal in about 15% (1). However, only 1 of the 199 babies with documented bacterial meningitis had completely normal CSF examinations at the time of initial evaluation. Gram stain of the spinal fluid should be positive in approximately 80% of infants with proven bacterial meningitis.

A number of adjuncts to diagnosis have evolved over the last several years. The most specific of these is the counterimmunoelectrophoresis (CIE) (38) test, utilized to detect specific bacterial antigens in body fluids. It has been reported to be positive in as many as two thirds of infants with *E. coli* meningitis, but it is detectable only when organisms are viable and found in culture.

Prognosis

Mortality rates of neonatal meningitis vary considerably among institutions and within the same institution at different periods of time. The overall mortality rate of the collaborative study (30%) is lower than previous reports. This suggests that supportive measures and changes in overall care of infants have improved over the last decade.

Long term sequelae include the entire spectrum of neurological and developmental compromise. At least 50% of survivors have demonstrable damage. Long term follow-up is necessary so that problems can be accurately appraised and, when possible, attended to by appropriate supportive measures, special schooling, and/or counseling of the parent and child.

THERAPY OF GRAM-NEGATIVE ENTERIC MENINGITIS OF INFANCY

Therapy is aimed at sterilization of the CSF. A collaborative study (1) was organized to examine the effects on infants of parenterally administered drugs as compared to parenterally administered antibiotics plus intrathecal gentamicin administered in the lumbar space. In this study, the CSF was sterilized 3–4 days after therapy began; no difference was seen in the group receiving intrathecal gentamicin. This study established that 1 mg of gentamicin administered daily into the lumbar intrathecal space for 3 days or longer had no discernible effect on the outcome of neonatal meningitis. The overall mortality in this study was 32% (37 infants).

The concentration of gentamicin in the CSF was 18–40 μg/ml for several hours after intrathecal administration (1). When examination of the spinal fluid was repeated 18–24 hr later, the concentration of the drug was 1–3 μg/ml which was similar to that measured in the CSF of infants receiving only parenteral therapy. The half-life of gentamicin in the CSF is approximately 5 hr (39). The aminoglycoside failed to enter the cerebral ventricles and was poorly distributed to other parts of the CSF space after administration in the lumbar intrathecal space.

Although the presence of ventriculitis has been documented in the majority of infants with fatal infections, the number of infants with ventriculitis at the time of diagnosis of meningitis is presently unknown. It is conceivable that the eradication of organisms from the ventricular space may be more rapidly achieved by direct instillation of drugs into that space. Therefore, the second phase of the collaborative neonatal meningitis study is presently attempting to ascertain whether such a technique will further reduce the morbidity and mortality from this infection. Of necessity, this study will evaluate any potentially harmful effects of multiple direct instillations of drugs into the ventricles.

Presently, recommended therapy consists of the administration of ampicillin, 100–200 mg/kg per day in 2–3 doses for 21 or more days. In addition, 15–30 mg/kg per day of kanamycin or 5–7.5 mg/kg per day of gentamicin in 2–3 doses for 21 or more days is also administrered to the infant with gram-negative meningitis. Intrathecal gentamicin in a dosage of 2–3 mg daily may be used in some neonates. In particular, if the CSF cultures remain positive, ventriculitis may be present and instillation of drug into the ventricle is indicated. Until the results of the collaborative study are known, routine tapping of the ventricles at the time of diagnosis is not recommended; instillation of aminoglycosides into the ventricles requires constant monitoring of the amount of drug present to avoid toxic accumulation.

SUMMARY

It is clear from the preceding material that bacterial infection is a significant threat to the newborn infant. The most common gram-negative organism causing meningitis in the neonate is *E. coli*, and the invasive capacity of this organism is correlated with its polysaccharide antigen, K_1. Careful studies of supportive care and antimicrobial therapy of these infections are continuing. If antibodies could be generated which protect the infant, it would be theoretically possible to guard against invasive disease by conferring antibody positivity to the mother and hence to the infant. Investigations along these lines are continuing.

Literature Cited

1. McCracken, G. H., Mize, G. 1976. A controlled study of intrathecal antibiotic therapy in gram-negative enteric meningitis of infancy. Report of the neonatal meningitis cooperative study group. *J. Pediatr.* 89:66–72
2. Overall, J. C. 1970. Neonatal bacterial meningitis. Analysis of predisposing factors and outcome compared with matched control subjects. *J. Pediatr.* 76:499–511
3. Smith, E. S. 1954. Purulent meningitis in infants and children. A review of 409 cases. *J. Pediatr.* 45:425–36
4. Wilfert, C. M. 1975. The neonate and gram negative bacterial infections. In *Progress in Clinical and Biological Research 3: Infections of the Fetus and the Newborn Infant,* ed. S. Krugman, A. G. Gershon, pp. 167–81. New York: Liss
5. Michael, J. G., Rosen, F. S. 1963. Association of "natural" antibodies to gram negative bacteria with the gamma 1 macroglobulins. *J. Exp. Med.* 118:619–23
6. Fireman, P., Zuchowski, D. A., Taylor, P. M. 1969. Development of human complement system. *J. Immunol.* 103:25–31
7. Propp, R., Alper, C. A. 1968. C3 synthesis in the human fetus and lack of transplacental passage. *Science* 162:672–73
8. Stossel, T. P., Alper, C. A., Rosen, F. S. 1973. Opsonic activity in the newborn. Role of properdin. *Pediatrics* 52:134–37
9. Miller, M. E. 1971. Demonstration and replacement of a functional defect of the 5th component of complement in newborn serum. A major tool in the therapy of neonatal septicemia. *Pediatr. Res.* 5:379–80 (Abstr.)
10. Miller, M. E. 1971. Chemotactic function in the human neonate. Humoral and cellular aspects. *Pediatr. Res.* 5:487–92
11. Miller, M. E. 1969. Phagocytosis in

the newborn infant. Humoral and cellular factors. *J. Pediatr.* 74:255–59

12. Forman, M. L., Stiehm, E. R. 1969. Impaired opsonic activity but normal phagocytosis in low birth weight infants. *N. Engl. J. Med.* 281:926–31
13. Tomasi, T. B., Tan, E. M., Solomon, A., Prendergast, R. A. 1965. Characteristics of an immune system common to external secretions. *J. Exp. Med.* 121:101–24
14. Amman, A. J., Stiehm, E. R. 1966. Immune globulin levels in colostrum and breast milk, and serum from formula and breast fed newborns. *Proc. Soc. Exp. Biol. Med.* 122:1098–1100
15. Sussman, S. 1961. The passive transfer of antibodies to *E. coli* 0111 B4 from mother to offspring. *Pediatrics* 27:308–13
16. Smith, C. W., Goldman, A. S. 1968. The cells of human colostrum. I. In vitro studies of morphology and function. *Pediatr. Res.* 2:103–9
17. Lascelleo, A. K., Gurner, B. W., Coombs, R. R. A. 1969. Some properties of human colostrol cells. *Aust. J. Exp. Biol. Med. Sci.* 47:349–60
18. Smith, C. W., Goldman, A. S. 1970. Interactions of lymphs and macrophages from human colostrum: Characteristics of the interacting lymphocytes, *J. Reticuloendothelial. Soc.* 8:91–104
19. Gyorgy, P. 1953. A hitherto unrecognized biochemical difference between human milk and cows milk. *Pediatrics* 11:98–108
20. Gyorgy, P. 1971. The uniqueness of human milk. Biochemical aspects. *Am. J. Clin. Nutr.* 24:970–75
21. Jollés, P., Jollés, J. 1961. Lysozyme from human milk. *Nature* 192:1187–88
22. Chandan, R. C., Shahani, K. M., Holly, R. G. 1964. Lysozyme content of human milk. *Nature* 204:76–77
23. Johansson, B. G. 1960. Isolation of an iron containing red protein from human milk. *Acta Chem. Scand.* 14:510–12
24. Masoon, P. L., Heremans, J. F., Dive, C. H. 1966. An iron binding protein common to many external secretions. *Clin. Chim. Acta* 14:735–39
25. Bullen, J. J., Rogers, H. J., Leigh, L. 1972. Iron binding proteins in milk and resistancc to *E. coli* infection in infants. *Br. Med. J.* Jan. 8, pp. 69–75
26. Kauffman, F., Vahine, G. 1945. Über die Bedeutung des serologeschen formenwechsels für die baktereophagen—wirkung in der coli—Gruppe. *Acta Pathol. Microbiol. Scand.* 22:119–37
27. Smith, D. E. 1927. Studies on pathogenic *E. coli* from bovine sources. IV. A biochemical study of the capsular substance. *J. Exp. Med.* 46:155–66
28. Grados, O., Ewing, W. H. 1970. Antigenic relationship between *Escherichia coli* and *Neisseria meningitidis. J. Infect. Dis.* 122:100–3
29. Kasper, D. L., Winkelhake, J. L., Zollinger, W. D., Brandt, B. L., Artenstein, M. S. 1973. Immunochemical similarity between polysaccharide antigens of *Escherichia coli* 07:K1 (L): NM and Group B *Neisseria meningitidis. J. Immunol.* 110:262–68
30. Barry, G. T. 1958. Colominic acid, a polymer of N-acetylneuraminic acid. *J. Exp. Med.* 107:507–21
31. Wolberg, G., DeWitt, C. W. 1969. Mouse virulence of K (L) antigen containing strains of *Escherichia coli. J. Bacteriol.* 100:730–37
32. Robbins, J. B., McCracken, G. H., Gotschlick, E. C., Ørskov, F., Ørskov, I., Hanson, L. A. 1974. Escherichia coli K1 capsular polysaccharide associated with neonatal meningitis. *N. Engl. J. Med.* 290:1216–20
33. Sarff, L. D., McCracken, G. H., Schiffer, M. S., Glode, M. P., Robbins, J. B., Ørskov, I, Ørskov, F. 1975. Epidemiology of *Escherichia coli* K1 in healthy and diseased newborns. *Lancet* 2:Pt 2; pp. 1099–1104
34. McCracken, G. H., Sarff, L. D., Glode, M. P., Mize, S. G., Schiffer, M. S., Robbins, J. B., Gotschlich, E. C., Ørskov, I., Ørskov, F. 1974. Relation between *Escherichia coli* K1 capsular polysaccharide antigen and clinical outcome in neonatal meningitis. *Lancet* 2: Pt. 1; pp. 246–50
35. Schiffer, M. S., Olveira, E., Glode,

M. P., McCracken, G. H., Sarff, L. M., Robbins, J. B. 1976. A review relation between invasiveness and the K1 capsular polysaccharide of *E. coli*. *Pediatr. Res.* 10:82–87

36. Naidoo, B. T. 1968. The cerebrospinal fluid in the healthy newborn infant. *S. Afr. Med. J.* 42:933–35

37. Bauer, C. H., New, M. I., Miller, J. N. 1965. Cerebrospinal fluid protein values of premature infants. *J. Pediatr.* 66:1017–22

38. Feigin, R. D., Wong, M., Shackelford, P., Stechenberg, B. W., Dunkle, L. M., Kaplan, S. 1976. Countercurrent immunoelectrophoresis of urine as well as of CSF and blood for diagnosis of bacterial meningitis. *J. Pediatr.* 89:773–75

39. Kaiser, A. B., McGee, Z. A. 1975. Aminoglycoside therapy of gram negative bacillary meningitis. *N. Engl. J. Med.* 293:1215–20

Ann. Rev. Med. 1978. 29:137–48
Copyright © 1978 by Annual Reviews Inc. All rights reserved

ANOREXIA NERVOSA: RECENT INVESTIGATIONS

◆7273

Katherine A. Halmi, M.D.
Department of Psychiatry, University of Iowa, Iowa City, Iowa 52242

"I do not remember that I did ever in all my practice see one, that was conversant with the Living so much wasted with the greatest degree of Consumption (like a Skeleton only clad with Skin)." This description by Richard Morton (1689) in his monograph *Phthisiologia* (1) is most likely the first recorded recognition of anorexia nervosa as a distinct illness. This malady, which occurs predominantly in pubertal females and young women, was regarded as a form of willful starvation resulting in emaciation with no other associated medical or psychiatric etiology by Sir William Gull (1874) who created its name, anorexia nervosa (2). Concurrently, this disorder, characterized by behavior directed towards losing weight, peculiar patterns of handling foods, weight loss, intense fear of gaining weight, disturbance of body image, and in women, amenorrhea, was described by Lasègue in 1873 (3). After Simmonds reported a "destroyed" pituitary gland in a woman who died of emaciation, anorexia nervosa was often erroneously diagnosed as Simmonds disease (4). Controversies over the psychological or physiological origin of this illness continue today. In the past 15 years, numerous investigations have been directed toward testing hypotheses concerning the etiology of anorexia nervosa.

TYPOLOGICAL DIAGNOSIS OF ANOREXIA NERVOSA

Many careful and comprehensive descriptions of the clinical characteristics of this disorder (5–12), a few large demographic surveys (13–15), and several follow-up studies (13, 16–20), have provided sufficient evidence for the operational criteria for anorexia nervosa (Table 1) and thus the necessary documentation that it is a disorder sui generis. The essential features of the illness are the behavior directed toward losing weight and the intense fear of gaining weight. Weight loss is accomplished by several means: (*a*) a drastic reduction in the total food intake with a disproportionate decrease in high carbohydrate- and fat-containing foods, (*b*) self-induced vomiting which often occurs after a bulimic episode, a carbohydrate-eating binge, (*c*) laxative and/or diuretic abuse, (*d*)

0066-4219/78/0401-0137$01.00

overactivity such as prolonged exercising, extensive cycling, or constant walking which will continue despite fatigue and weakness.

The patients have an intense fear of gaining weight that extends to a fear of losing control and becoming obese. However, this does not cause them to avoid food. While limiting themselves to a narrow selection of low-calorie foods, they often engage in elaborate meal preparations for others. They frequently collect recipes and are constantly preoccupied with thoughts of food and its calorie content. Hoarding, concealing, crumbling, or throwing food away are patterns of food handling frequently observed. The term *anorexia* is actually a misnomer since loss of appetite is usually rare until late in the illness. Some patients cannot exert continuous control over their voluntary restriction of food intake and will have bulimic episodes.

In addition to the patients' intense fear of gaining weight, they manifest a distorted body image which expresses itself as a conviction that they look well, or even overweight, in the face of increasing cachexia. They are preoccupied with their body size and often gaze at their image in the mirror. Most patients steadfastly deny the illness and are disinterested in, even resistive to, therapy. Many of the adolescents have a delayed psychosocial sexual development, and adults have a markedly decreased interest in sex with the onset of the illness. Compulsive behavior, such as hand washing, may be present during the illness.

The patient usually comes to medical attention when weight loss is observable. As the loss becomes profound, physical signs such as hypothermia (as low as 35°C), dependent edema, bradycardia, hypotension, lanugo, and a variety of metabolic changes occur. The females have amenorrhea, which often appears before noticeable weight loss has occurred. A higher-than-expected frequency of urogenital abnormalities and Turner's Syndrome has been found in patients with anorexia nervosa (21–23).

This disorder occurs predominantly in females. Only 4–6% of the patients are males (15). The disorder may be as prevalent as 1 in 200 females in the high-risk age group between 12 and 18 years (24). The age of onset is usually early to late adolescence, although it can range from prepuberty to the early 30s (rare). The course may be unremitting until death. It may fluctuate with several ex-

Table 1 Operational criteria for anorexia nervosa

A. Refusal to maintain body weight over a minimal normal weight for age and height

B. Weight loss of at least 25% of original body weight, or, if under 18 years of age, weight loss from original body weight plus projected weight gain expected on pediatric growth charts may be combined to make the 25%

C. Disturbance of body image with inability to accurately perceive body size

D. Intense fear of becoming obese. This fear does not diminish as weight loss progresses

E. No known medical illness that would account for loss.

F. Amenorrhea (in females)

acerbations and remissions or occur as a single episode, but the last is most common. The most tragic complication is death through starvation. Follow-up studies involving a large series of patients have had mortality rates between 15 and 21% (13, 16–19). Sisters of anorectic patients are at an increased risk for developing the illness and a greater-than-expected occurrence of this disorder is present in the parents of the patients (13, 25). Adequate studies have not been conducted to establish definite predisposing factors. However, there is evidence that in some patients the onset of illness is associated with a stressful life situation (15). There is also some evidence that patients tend to be very well behaved and perfectionistic "model children" (18, 26).

In diagnosing anorexia nervosa, a differentiation must be made from other psychiatric disorders such as depression, schizophrenia, and hysteria. In these, weight loss, peculiar eating behavior, and vomiting can occur; however, the full criteria for anorexia nervosa will not be met. It must be ascertained that no known medical illness can account for the weight loss. Anorexia nervosa must also be differentiated from bulimia in which patients have bulimic episodes and self-induced vomiting while maintaining their weight within a normal range.

METABOLIC STUDIES

Most of the metabolic changes found in anorexia nervosa are also present in starvation states secondary to other causes, and thus do not provide etiologic clues for this malady. Although these metabolic aberrations revert to normal with nutritional rehabilitation, some of the initial abnormalities cannot be explained by a simple starvation effect.

Abnormalities in hematopoiesis were reported as early as 1959 by Carryer, Berkman & Mason (27). They found a leukopenia and relative lymphocytosis in their acutely emaciated patients. Both Silverman (11) and Lampert & Lau (28) reported hypoplasia of bone marrow in starved anorectics. The bone marrow alterations reverted to normal after weight gain. More specifically, Mant & Faragher (29) described the bone marrow abnormalities as hypocellularity and noted the presence of a large amount of gelatinous acid mucopolysaccharide. They found this to reverse with clinical improvement and felt it represented a physiological response to starvation. They present evidence that these changes are of minimal functional significance in most patients and postulate that depletion of fat or the presence of background acid mucopolysaccharide would adversely affect the local environment for normal hematopoiesis.

Low erythrocyte sedimentation rates and plasma fibrinogen levels were measured during the emaciated state of anorexia nervosa. These levels were found to increase with weight recovery (30). The initial hypofibrinogenemia was thought to be caused by decreased production of fibrinogen by the liver responding to the malnourised state. In a study of the fibrinolytic enzyme system in anorexia nervosa, Ogston & Ogston (31) found that the high plasminogen activator levels present in patients with untreated anorexia reverted to the range of control subjects after weight gain. They also found the levels of plasminogen, fibrin degradation products, α-1-antitrypsin, and α-2-macroglobulin in emaciated anorectics did not differ from those in normal women.

Most studies have reported total plasma proteins in anorectic patients to be within normal range (11, 32). In one study (32), the percentages of the α-2-fraction of the serum proteins, albumin, and γ-globulin were unchanged, but the β-globulin percentage rose significantly with nutritional rehabilitation in anorectic patients. Russell (33) reported that plasma aminograms (18 amino acids analyzed) were normal in 19 anorexia nervosa patients, in contrast with the pattern of reduced essential amino acids and increased unessential amino acids seen in kwashiorkor. Although serum albumin levels are normal (11, 32, 33), Russell calculated the rate of degradation of labelled I^{131}-albumin in 15 anorectics and found a small but significant reduction in the breakdown rate in the anorectic patients compared to normal subjects (33).

The first association of carotenemia with anorexia nervosa was made by Dally in 1959 (34). He reported that an anorectic patient who had eaten large amounts of fruit and cooked green vegetables (no carrots) for ten years had a serum carotene level of 240 μg/ml (normal, 50–100 μg/ml) and a serum vitamin A concentration of 182 IU/100 ml (normal, 80–130 IU/100 ml). Pops & Schwabe (35) found hypercarotenemia in 9 of 12 anorexia nervosa patients. They hypothesized a decreased catabolism of β-lipoprotein, the major carrier of carotene in plasma. Most recently, significant elevations of serum β-carotene, vitamin A esters, vitamin A alcohol, and vitamin A acid were found in anorexia nervosa patients compared to control groups of normal subjects, patients with cachexia from other causes, patients with ovarian failure, and patients with pituitary suppression (36). There was no significant correlation of hypercarotenemia with elevated serum cholesterol levels. It is of interest that Crisp (37) reported only 1 anorectic with persistently elevated cholesterol levels in a group of 11 anorexia nervosa patients with hypercarotenemia. Serum lipid electrophoresis has been reported normal in several studies (36, 38). Robboy, Sato & Schwabe (36) are now hypothesizing that an acquired but reversible error of metabolism of β-carotene and vitamin A precursors occurs in anorexia nervosa. An alternative and equally plausible explanation is that a reliable history of dietary intake is often unobtainable.

The relationship of cholesterol and lipid metabolism to anorexia nervosa is still an enigma. Kleinfelter (39) first reported higher-than-normal serum cholesterol levels in the younger age anorectic groups. This difference was no longer present in the older patients. Subsequently, Crisp, Blendis & Pawan (40) compared serum cholesterol levels in 37 female anorectics with 37 normal age- and sex-matched controls. Although the cholesterol levels were higher in the anorectics, there was a wide scatter of levels in the patient and control groups. Serum cholesterol levels were unrelated to protein-bound iodine levels, 24-hr thyroid uptake of I^{131}, and basal metabolic rate. Lack of correlation of serum cholesterol levels to thyroid function in anorexia nervosa was confirmed by another study (32). Halmi & Fry (41) were not able to prove the hypothesis of

Blendis & Crisp (42) that high serum cholesterol levels occurred only in those anorectics who had carbohydrate-eating binges. In a more recent study, Nestel (43) found in anorectic girls with hypercholesterolemia that the daily intake of cholesterol exceeded the total fecal excretion of steroids, which indicates that dietary cholesterol was being retained. Bile acid excretion was markedly reduced in those patients. The author concluded the hypercholesterolemia most likely reflected the hypocaloric state, since the serum cholesterol has been shown to rise in normal subjects who have reduced their caloric intake.

Some of the metabolic changes seen in anorexia nervosa are directly related to the behavior of the patients. There are anorectic patients who engage in self-induced vomiting or abuse purgatives and diuretics. Since these patients will often deny the above behaviors, extensive and unnecessary investigations for hypokalemic alkalosis are made (44–46). Of 20 anorectic patients with hypokalemia reported by Wigly (47), 11 showed evidence of renal disorder. Autopsy findings on one of these patients showed renal tubular vacuolation.

Acute gastric dilatation is a rare complication of anorexia nervosa, and, in the cases reported (48–50), the pathogenesis was not determined. Recently aortography done on an anorectic patient with acute gastric dilatation showed a "superior mesenteric artery syndrome" (51). The superior mesenteric artery crosses in front of the duodenum which it may compress, causing symptoms of partial obstruction. It is conceivable that severe weight loss, whether in anorexia nervosa or in other malnourished states, will cause a change in the amount of fat in the mesentery and subsequently affect the position of the superior mesentery artery and its branches.

All of the metabolic changes discussed above have been attributed to the severe reduction in caloric intake and the emaciated state of the patient. Electrocardiographic changes such as flattening or inversion of the T wave, ST segment depression, and lengthening of the QT interval have been noted in the emaciated state of anorexia nervosa and will revert to normal with nutritional rehabilitation (11, 52). Thurston & Marks (52) noted that these electrocardiographic changes are similar to those frequently observed in patients with head injuries. In acute intracranial disease, sympathetic centers are stimulated in the hypothalamus and this results in release of catecholamines within the myocardium. The authors postulate a similar phenomenon occurs in anorexia nervosa and that these catecholamines may cause damage to the contractile apparatus and to the cell membrane of the myocardial cells, resulting in the electrocardiographic abnormalities described.

ENDOCRINE INVESTIGATIONS

With the development of more accurate measurements for hypothalamic and pituitary hormones, a resurgence of interest has emerged for reexamining endocrine function in anorexia nervosa. Past endocrine studies are marred by the use of vague and inconsistent criteria for the diagnosis of the illness, and because many studies were done while the patients were being treated with phenothiazines and other drugs that in themselves affect endocrine and hypothalamic function. Recent studies of thyroid function in anorexia nervosa have shown that the mean basal thyrotropin (TSH) level and the mean maximum increase in TSH after administration of thyrotropin-releasing hormone (TRH) were comparable to those in normal subjects (53). Thyroxin (T_4) levels are consistently found to be in the low normal range and serum triiodothyronine (T_3) levels are decreased (53–55). Similar findings are present in other malnourished states. Because TSH reserve is within a normal range in emaciated states, it has been postulated that a defect in hypothalamic function leads to decreased TRH. This, however, would not be a defect specific to anorexia nervosa but rather a characteristic of any emaciated state (55–57). Since plasma free T_4 levels are within the normal range for the emaciated state in anorexia nervosa and plasma free T_3 levels are markedly reduced (55), and since the same phenomenon has been observed in patients with protein calorie malnutrition (57), it has been postulated that there is a decrease of peripheral T_4-to-T_3 conversion in emaciated states. Elevated serum concentrations of 3,3′,5′-triiodothyronine (reverse T_3, or rT_3) and low levels of serum free T_3 were present in patients with protein calorie malnutrition and in obese patients fasting on no-carbohydrate hypocaloric diets (58–60). Thus, it seems that dietary carbohydrate is an important regulator of T_3 production in man. In a recent study, 9 of 16 anorexia nervosa patients had slightly elevated serum rT_3 levels and low serum T_3 levels compared to control subjects (61). To date, only teleological explanations have been given for these changes, which are not specific to anorexia nervosa.

It is of interest that short-term administration of T_3 reversed the cortisol metabolic abnormalities such as (*a*) increased mean 24-hr plasma cortisol concentration, (*b*) increased cortisol half-life, and (*c*) increased tetrahydrocortisol/tetrahydrocortisone ratio in emaciated anorexia nervosa patients (62). The investigators of this study have proposed that the avoidance of food and insidious weight loss lead to subnormal plasma triiodothyronine concentration and decreased 11-hydroxysteroid-dehydrogenase and 5-α-reductase enzyme activity; this then results in hypothyroid-like abnormalities of cortisol and androgen metabolism (63). They also state the changes in thyroxin and steroid metabolism may reflect hepatic dysfunction in the face of malnutrition. The reduced metabolic clearance rate of cortisol and the incomplete suppression of ACTH and cortisol levels by dexamethasone have been reported in patients with protein calorie malnutrition as well as in anorexia nervosa (64, 65).

Since amenorrhea is a major sign of anorexia nervosa, extensive studies have been conducted on the hypothalamic-pituitary-ovarian axis. One study, in which measurements of plasma luteinizing hormone (LH) concentrations were made at 20-min intervals for 24-hr periods, showed that women in the starvation stage of anorexia nervosa had a LH secretion pattern that resembled the pattern found in prepubertal and pubertal children (66). This regressed pattern of LH secretion changed to a more "mature" secretion pattern with weight gain. The regression to a prepubertal pattern of hormone secretion was also shown in another study in which the follicle-stimulating hormone (FSH) response to gonadotropin-releasing hormone (GnRH) was markedly greater than the LH response in emaciated anorectic women. This is not unlike the pattern reported in prepubertal girls (67). (The reversion to a more mature pattern of response to GnRH after weight gain gives some support to the hypothesis that sexual maturation is related to the attainment of a critical body mass.) Although there has been some variability, most studies have reported low serum LH and estrogen levels during the emaciated state of this illness (68–70). Most consistently, these studies have shown an impaired LH response to GnRH during the acute state of emaciation, with an increased response after weight gain (68, 70, 71). In one study, responses of LH and FSH to GnRH were quantitatively normal but abnormally delayed (69). Two studies have shown that in the malnourished state the response of LH to clomiphene was either absent or incomplete (72, 73). After weight gain the patients showed an initial rise in LH after clomiphene, but often the second LH peak and subsequent menstruation did not occur. Thus, in spite of the fact that restoration of a normal body weight is prerequisite to resumption of menstruation in this disorder, factors other than the nutritional state (most likely psychological factors) contribute to the prolonged amenorrhea and, thus, limit the effectiveness of clomiphene treatment. In another study, administration of GnRH (500 μg) parenterally 3 times a day over a 4-week period induced follicular maturation and ovulation (74, 75). It should be noted during the period of GnRH administration there was no change in typical anorectic behavior or weight.

Much emphasis has been placed on elevated basal growth hormone (GH) levels in anorexia nervosa (32, 76, 77). A careful examination of these studies reveals that fewer than one third of the patients had elevated basal GH levels; in fact, the remarkable finding is the extreme variability of the fasting GH levels. The increase in serum GH has been attributed to caloric deprivation seen in protein calorie malnutrition. Studies have shown that in order for serum GH to rise in protein calorie malnutrition, severe protein deprivation was necessary and was associated with a drop in serum alanine and albumin levels (78, 79); however, these levels are normal in anorexia nervosa (38). The response of GH to various stimulation tests has also been extremely variable. In one study, only 8 of 22 patients responded to 0.5 g of L-dopa p.o. with an increment in serum GH concentration greater than 6 ng/ml. After nutritional rehabilitation, only 6 of 22 patients had a normal response to L-dopa (80). In this latter study, 7 patients had the L-dopa test repeated after 2 days of diethylstilbesterol (0.5 mg per day), with no significant enhancement of the GH response to L-dopa. Of the 12 patients tested with 0.75 mg of apomorphine, a dopamine receptor agonist given subcutaneously, only 4 patients responded with an increase in GH secretion (80). This suggests there may be a specific abnormality in dopaminergic regulation of GH secretion in anorexia nervosa. Two other studies have shown a blunted GH response to insulin-induced hypoglycemia in emaciated anorexia patients (81, 82). In these studies, however, the stimulation tests were not repeated after the patients gained weight; thus, it is not known if the abnormal growth hormone responses would have been retained as in the case with the L-dopa and apomorphine stimulation tests. In one 11-yr-old anorectic patient, failure to grow and deficiency in GH was documented. After four years of treatment and nutritional rehabilitation, GH responses to provocative tests returned to normal and a 24-hr integrated GH concentration was also normal for age. During this time, the patient grew 14 cm (83). Normal regulation of gonadotropin and GH secretion is exceedingly complex, and the precise role of biogenic amines in the regulation of these hormones, and in feeding behavior as well, is not clear. Mean basal prolactin levels do not differ from those measured in normal controls (84, 85). The adenohypophysis of an anorexia nervosa patient was examined with a battery of differential stains and immunocytochemically (86). The ACTH, GH, TSH, and prolactin cells stained with the same intensity as those in normal control pituitaries. The gonadotrophs immunostained for LH stained much more faintly in the anorectic's pituitary than in normal glands. These findings reflect a selective pituitary deficiency in the emaciated state of anorexia nervosa.

There is compelling evidence that the noradrenergic neurotransmitter system is eminently involved both in the control of feeding behavior (87–89) and in primary depression (90–92). A recent study has shown a relationship to exist between urinary secretion of 3-methoxy,4-hydroxy-phenylglycol (MHPG), a major metabolite of brain norepinephrine, and the symptom of depression in anorexia nervosa patients (93). Urinary MHPG was significantly lower in acutely ill anorectic patients than it was in female controls. An increase in urinary MHPG significantly correlated with a decrease in depression ratings after weight gain. It is of special importance that the change in urinary MHPG over treatment was not significantly associated with the change in patients' weight, activity level, urinary creatinine, or other urinary catecholamines. This study provided the first direct measurement of catecholamine metabolism in anorexia nervosa and suggests that the noradrenergic transmitter system may be involved in the pathogenesis of this disorder.

CONCLUSION

In this paper, I have: 1. described the typological diagnosis of anorexia nervosa, 2. discussed recent hematological, lipid, electrolyte, gastrointestinal, and electrocardiographic studies in anorexia nervosa, and 3. reviewed recent endocrine investigations including thyroid, adrenal cortical, gonadotropin, growth hormone, and catecholamine studies in anorexia nervosa. Although all of the reviewed studies have contributed greatly to the formulation of descriptive criteria for anorexia nervosa, they have not supported any single hypothesis concerning the etiology of anorexia nervosa. Further studies are needed to ascertain relationships between the typical anorectic behavior and the endocrine and catecholamine changes observed in this disorder.

Literature Cited

1. Morton, R. 1720. *Phthisiologia—Or a Treatise of Consumption*. London: Smith. 2nd ed.
2. Gull, W. 1874. Anorexia Nervosa (Apepsia hysterica, anorexia hysterica). *Trans. Clin. Soc. London*, 7:22–28
3. Lasègue, E. P. 1873. De L'Anorexie Hysterique. *Arch. Gen. Med.* 1:385
4. Sheehan, H. L., Summers, B. K. 1949. The syndrome of hypopituitarism. *Q. J. Med.* 18:319
5. King, A. 1963. Primary and secondary anorexia nervosa syndrome. *Br. J. Psychiatry* 109:470–79
6. Sours, J. A. 1968. Clinical studies in anorexia nervosa syndrome. *N. Y. State J. Med.* 68:1363–69
7. Crisp, A. H. 1974. Primary anorexia nervosa or adolescent weight phobia. *Practitioner* 212:525–35
8. Burch, H. 1966. Anorexia nervosa and its differential diagnosis. *J. Nerv. Ment. Dis.* 141:555–66
9. Galdston, R. 1974. Mind over matter. *J. Child Psychol. Psychiatry Allied Discip.* 13:246–63
10. Halmi, K., Brodland, G., Loney, J. 1973. Prognosis in anorexia nervosa. *Ann. Intern. Med.* 78:907–9
11. Silverman, J. A. 1974. Anorexia nervosa: clinical observations in a successful treatment plan. *J. Pediatr.* 84:68–73
12. Smart, D. E., Beumont, P. J., George, G. C. 1976. Some personality characteristics of patients with anorexia nervosa. *Br. J. Psychiatry* 128:57–60
13. Theander, S. 1970. Anorexia nervosa. *Acta Psychiatr. Scand. Suppl.* 214
14. Dally, P. 1969. *Anorexia Nervosa*. New York: Grune & Stratton
15. Halmi, K. 1974. Anorexia nervosa: demographic and clinical features in 94 cases. *Psychosom. Med.* 36:18–26
16. Halmi, K. A., Brodland, G., Rigas, C. 1975. A follow-up study of 79 patients with anorexia nervosa: an evaluation of prognostic factors and diagnostic criteria. *Life Hist. Res. Psychopathol.* 4:290–301
17. Kay, D. W. K., Leigh, D. 1954. The natural history, treatment and prognosis of anorexia nervosa, based on a study of 38 patients. *J. Ment. Sci.* 100:411–30
18. Crisp, A. H. 1965. Clinical and therapeutic aspects of anorexia nervosa; a study of 30 cases. *J. Psychosom. Res.* 9:67–78
19. Seidensticker, J. F., Tzagournis, M. 1968. Anorexia nervosa—clinical features and long term follow up. *J. Chronic Dis.* 21:361–67
20. Morgan, H. G., Russell, G. F. M. Value of family background in clinical features as predictors of long term outcome in anorexia nervosa: four-year follow up study of 41 patients. *Psychol. Med.* 5:355–71
21. Halmi, K., Rigas, C. 1973. Urogenital malformations associated with anorexia nervosa. *Br. J. Psychiatry* 122:79–82
22. Liston, E. H., Shershow, L. W. 1973. Concurrence of anorexia nervosa and gonadal dysgenesis. *Arch. Gen. Psychiatry* 29:834–36
23. Kron, L., Katz, J. L., Gorzynsky, G., Weiner, H. 1977. Anorexia nervosa and gonadal dysgenesis. *Arch. Gen. Psychiatry* 34:332–35
24. Crisp, A. H., Palmer, R. L., Kalucy, R. S. 1976. *Br. J. Psychiatry* 128:549–54
25. Kalucy, R. S., Crisp, A. H., Harding, B. 1975. *A study of 56 families with anorexia nervosa*. Presented at Int. Colloq. Psychosom. Med., 2nd Congr., Rome
26. Halmi, K. A., Goldberg, S., Eckert, E., Casper, R., Davis, J. 1977. Pretreatment evaluation in anorexia nervosa. In *Anorexia Nervosa*, ed. R. A. Vigersky, pp. 43–54. New York: Raven
27. Carryer, H. J., Berkman, J. M., Mason, H. L. 1959. Relative lymphocytosis in anorexia nervosa. *Staff Meet. Mayo Clinic* 34:426–32
28. Lampert, F., Lau, B. 1976. Bone marrow hypoplasia in anorexia nervosa. *Eur. J. Pediatr.* 124:65–71
29. Mant, M. J., Faragher, B. S. 1972. The haematology of anorexia nervosa. *Br. J. Haematol.* 27:737–49
30. Anyan, W. R. 1974. Changes in erythrocyte sedimentation rate and fibrinogen during anorexia nervosa. *J. Pediatr.* 85:525–27

31. Ogston, D., Ogston, W. D. 1976. The fibrinolytic enzyme system in anorexia nervosa. *Acta Haematol.* 55:230–33
32. Kanis, J. A., Brown, P., Fitzpatrick, K., Hibbert, D. J., Horn, D. B., Narin, I. M., Shirling, D., Strong, J. A., Walton, H. J. 1974. Anorexia nervosa: a clinical, psychiatric and laboratory study. *Q. J. Med.* 43:321–38
33. Russell, G. F. M. 1969. Metabolic, endocrine and psychiatric aspects of anorexia nervosa. *Sci. Basis Med. Ann. Rev.* 14:236–55
34. Dally, P. J. 1959. Carotenaemia occurring in a case of anorexia nervosa. *Br. Med. J.* 2:1333
35. Pops, M. A., Schwabe, A. D. 1968. Hypercarotenemia in anorexia nervosa. *J. Am. Med. Assoc.* 205:121–22
36. Robboy, M. S., Sato, A. S., Schwabe, A. D. 1974. The hypercarotenemia in anorexia nervosa: a comparison of vitamin A and carotene levels in various forms of menstrual dysfunction and cachexia. *Am. J. Clin. Nutr.* 27:326–67
37. Crisp, A. H., Stonehill, E. 1967. Hypercarotenemia as a symptom of weight phobia. *Postgrad. Med. J.* 43:721–23
38. Halmi, K. A., Struss, A., Stegink, L. Amino acids in anorexia nervosa. In preparation
39. Klinefelter, H. F. 1965. Hypercholesterolemia in anorexia nervosa. *J. Clin. Endocrinol. Metab.* 25:1520–21
40. Crisp, A. H., Blendis, L. M., Pawan, G. L. S. 1968. Aspects of fat metabolism in anorexia nervosa. *Metabolism* 17:1109–18
41. Halmi, K. A., Fry, M. 1974. Serum lipids in anorexia nervosa. *Biol. Psychiatry* 8:159–67
42. Blendis, L. M., Crisp, A. H. 1968. Serum cholesterol levels in anorexia nervosa. *Postgrad. Med. J.* 44:327–30
43. Nestel, P. J. 1974. Cholesterol metabolism in anorexia nervosa and hypercholesterolemia. *J. Clin. Endocrinol. Metab.* 38:325–28
44. Rodger, N. W., Collyer, J. A. 1970. Anorexia nervosa with concealed hyperphagia and self induced vomiting, hypokalemic alkalosis and normal aldosterone excretion. *Can. Med. Assoc. J.* 130:169–71
45. Love, D. R., Brown, J. J., Fraser, R., Lever, A. F., Robertson, J. I. S., Timbury, G. C., Thomson, S., Tree, M. 1971. An unusual case of self-induced electrolyte depletion. *Gut* 12:284–90
46. Wallace, M., Richards, P., Chesser, E., Wrong, O. 1968. Persistent alkalosis and hypokalemia caused by surreptitious vomiting. *Q. J. Med.* 37:577–88
47. Wigley, R. D. 1960. Potassium deficiency in anorexia nervosa, with reference to renal tubular vacuolation. *Br. Med. J.* 2:110–13
48. Jennings, K. P., Klidjian, A. M. 1974. Acute gastric dilatation in anorexia nervosa. *Br. Med. J.* 2:477–78
49. Russell, G. F. M. 1966. Acute dilatation of the stomach in a patient with anorexia nervosa. *Br. J. Psychiatry* 112:203–7
50. Evans, D. S. 1968. Acute dilatation and spontaneous rupture of the stomach. *Br. J. Surg.* 55:940–42
51. Halmi, K. A. 1977. The superior mesenteric artery syndrome in anorexia nervosa. Unpublished data
52. Thurston, J., Marks, P. 1974. Electrocardiographic abnormalities in patients with anorexia nervosa. *Br. Heart J.* 36:719–23
53. Moshang, T., Parks, J. S., Baker, L., Vaidya, V., Utiger, R. D., Bongiovanni, A. M., Snyder, P. J. 1975. Low serum triiodothyronine in patients with anorexia nervosa. *J. Clin. Endocrinol. Metab.* 40:470–73
54. Miyai, K., Yamamoto, T., Azukizawa, M., Ishibashi, K., Kumahara, Y. 1975. Serum thyroid hormones and thyrotropin in anorexia nervosa. *J. Clin. Endocrinol. Metab.* 40:334–38
55. Croxson, M. S., Ibbertson, H. K. 1977. Low serum triiodothyronine (T_3) and hypothyroidism in anorexia nervosa. *J. Clin. Endocrinol. Metab.* 44:167–74
56. Pimstone, B., Becker, D., Hendricks, S. 1973. TSH response to synthetic thyrotropin-releasing hormone in human protein-calorie malnutrition. *J. Clin. Endocrinol. Metab.* 36:779–83
57. Chopra, I. J., Smith, S. R. 1975.

Circulating thyroid hormones and thyrotropin in adult patients with protein-calorie malnutrition. *J. Clin. Endocrinol. Metab.* 40:221–27

58. Vagenakis, A. G., Burger, A., Portnay, G. I., Rudolph, M., O'Brian, J. T., Azizi, F., Arky, R. A., Nicod, P., Ingbar, S. H., Braverman, L. E. 1975. Diversion of peripheral thyroxine metabolism from activating to inactivating pathways during complete fasting. *J. Clin. Endocrinol. Metab.* 41:191–94
59. Chopra, I. J., Chopra, U., Smith, S. R., Reza, M., Solomon, D. H. 1975. Reciprocal changes in serum concentrations of 3,3',5'-triiodothyronine (reverse T_3) and 3,3'5-triiodothyronine (T_3) in systemic illnesses. *J. Clin. Endocrinol. Metab.* 41:1043–49
60. Spaulding, S. W., Chopra, I. J., Sherwin, R. S., Lyall, S. S. 1976. Effect of calorie restriction and dietary composition on serum T_3 and reverse T_3 in man. *J. Clin. Endocrinol. Metab.* 42:197–200
61. Burman, K. D., Vigersky, R. A., Loriaux, D. L., Strum, D., Djuh, Y., Wright, F. D., Wartofsky, L. 1977. Investigations concerning thyroxine deiodination pathways in patients with anorexia nervosa. See Ref. 26, pp. 255–62
62. Boyar, R. M., Hellman, L. D., Roffwarg, H., Katz, J., Zumoff, B., O'Connor, J., Bradlow, L., Fukushima, D. K. 1977. Cortisol secretion and metabolism in anorexia nervosa. *N. Engl. J. Med.* 296:190–93
63. Bradlow, H. L., Boyar, R. M., O'Connor, J., Zumoff, B., Hellman, L. 1976. Hypothyroid-like alterations in testosterone metabolism in anorexia nervosa. *J. Clin. Endocrinol. Metab.* 43:571–74
64. Takahara, J., Hosogi, H., Yunoki, S., Hashimoto, K., Uneki, T., Ofuju, T. 1976. Hypothalamic pituitary adrenal function in patients with anorexia nervosa. *Endocrinol. Jpn.* 23:451–56
65. Danowski, T. S., Livstone, E., Gonzales, A. R., Jung, Y., Khurana, R. C. 1972. Fractional and partial hypopituitarism in anorexia nervosa. *Hormones* 3:105–18
66. Boyar, R. M., Katz, J., Finkelstein, J. W., Kapen, S. Weiner, H., Weitzman, E. D., Hellman, L. 1974. Immaturity of the 24-hour luteinizing hormone secretory pattern. *N. Engl. J. Med.* 291:861–65
67. Sherman, B. M., Halmi, K. A., Zamudio, R. 1975. LH and FSH response to gonadotropin-releasing hormone in anorexia nervosa: effect of nutritional rehabilitation. *J. Clin. Endocrinol. Metab.* 41:135–42
68. Halmi, K. A., Sherman, B. M. 1975. Gonadotropin response to LH-RH in anorexia nervosa. *Arch. Gen. Psychiatry* 32:875–78
69. Vigersky, R. A., Loriaux, D. L., Andersen, A. E., Mecklenburg, R. S., Vaitukaitis, J. L. 1976. Delayed pituitary hormone response to LRF and TRF in patients with anorexia nervosa and with secondary amenorrhea associated with simple weight loss. *J. Clin. Endocrinol. Metab.* 43:893–900
70. Warren, M. P., Jewelewicz, R., Dyrenfurth, I., Ans, R., Khalaf, S., Vande Wiele, R. L. 1975. The significance of weight loss in the evaluation of pituitary response to LH-RH in women with secondary amenorrhea. *J. Clin. Endocrinol. Metab.* 40:601–11
71. Beumont, P. J., George, G. C., Pimstone, B. L., Vinik, A. I. 1976. Body weight and the pituitary response to hypothalamic releasing hormones in patients with anorexia nervosa. *J. Clin. Endocrinol. Metab.* 43:487–96
72. Aono, T., Kinugasa, T., Yamamoto, T., Miyake, A., Kurachi, K. 1975. Assessment of gonadotrophin secretion in women with anorexia nervosa. *Acta Endocrinol.* 80:630–41
73. Wakeling, A., Marshall, J. C., Beardwood, C. J., Souza, V. F., Russell, G. F. M. 1976. The effects of clomiphene citrate on the hypothalamic-pituitary-gonadal axis in anorexia nervosa. *Psychol. Med.* 6:371–80
74. Nillius, S. J., Fries, H., Wide, L. 1975. Successful induction of follicular maturation and ovulation by prolonged treatment with LH-releasing hormone in women with anorexia nervosa. *Am. J. Obstet. Gynecol.* 122:921–28
75. Nillius, S. J., Wide, L. 1975. Go-

nadotrophin-releasing hormone treatment for induction of follicular maturation and ovulation in amenorrhoeic women with anorexia nervosa. *Br. Med. J.* 3:405–8

76. Lundberg, P. O., Walinder, J., Werner, I., Wide, L. 1972. Effects of thyrotrophin-releasing hormone on plasma levels of TSH, FSH, LH and GH in anorexia nervosa. *Eur. J. Clin. Invest.* 2:150–53
77. Garfinkel, P. E., Brown, G. M., Stancer, H. C., Moldofsky, H. 1975. Hypothalamic-pituitary function in anorexia nervosa. *Arch. Gen. Psychiatry* 32:739–44
78. Pimstone, B. L., Barbezat, G., Hansen, J. P. L. 1968. Studies on growth hormone secretion in protein calorie malfunction. *Am. J. Clin. Nutr.* 21:482–85
79. Samuel, A. M., Deshpande, U. R. 1972. Growth hormone levels in protein calorie malnutrition. *J. Clin. Endocrinol. Metab.* 35:863–67
80. Sherman, B. M., Halmi, K. A. 1977. The effect of nutritional rehabilitation on hypothalamic-pituitary function in anorexia nervosa. See Ref. 26, pp, 211–24
81. Neri, V., Ambrosi, B., Beck-Peccoz, P., Travaglini, P., Faglia, G. 1972. Growth hormone regulation and hypothalamic-pituitary-adrenal function in anorexia nervosa. *Folia Endocrinol.* 25:143–51
82. Brauman, H., Gregoire, F. 1975. The growth hormone response to insulin induced hypoglycaemia in anorexia nervosa and control underweight or normal subjects. *Eur. J. Clin. Invest.* 5:289–95
83. Huseman, C., Johanson, A. 1975. Growth hormone deficiency in anorexia nervosa. *J. Pediatr.* 87:946–48
84. Beumont, P. J., Friesen, H. G., Gelder, M. G., Kolakowska, T. 1974. Plasma prolactin and luteinizing hormone levels in anorexia nervosa. *Psychol. Med.* 4:219–21
85. Hafner, R. J., Crisp, A. H., McNeilly, A. S. 1976. Prolactin and gonadotrophin activity in females treated for anorexia nervosa. *Postgrad. Med. J.* 52:76–79
86. Halmi, K. A. 1976. Selective pituitary deficiency in anorexia nervosa. In *Hormones, Behavior, and Psychopathology*, ed. E. J. Sachar, pp. 285–89. New York: Raven
87. Gold, R. M. 1973. Hypothalamic obesity: the myth of the ventromedial nucleus. *Science* 182:488–90
88. Margules, D. L., Lewis, M. J., Dragovich, J. A. 1972. Hypothalamic norepinephrine: circadian rhythms: the control of feeding behavior. *Science* 178:640–43
89. Leibowitz, S. F. 1970. Hypothalamic β-adrenergic "satiety" system antagonizes an α-adrenergic "hunger" system in the rat. *Nature* 226:963–64
90. Schildkraut, J. J., Keeler, B. A., Papousek, M., Hartmann, E. 1973. MHPG excretion in depressive disorders: relation to clinical subtypes and desynchronized sleep. *Science* 181:762–64
91. Maas, J. W., Fawcett, J. A., Dekirmenjian, H. 1972. Catecholamine metabolism, depressive illness and drug response. *Arch. Gen. Psychiatry* 25:252–62
92. Beckmann, H., Goodwin, F. K. 1975. Antidepressant response to tricyclis and urinary MHPG and unipolar patients. *Arch. Gen. Psychiatry* 32:17–21
93. Halmi, K. A., Dekirmenjian, H., Davis, J. M., Casper, R., Goldberg, S. Catecholamine metabolism in anorexia nervosa. *Arch. Gen. Psychiat.* In press

Ann. Rev. Med. 1978. 29:149–75
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CHEMOTHERAPY IN PREVENTION OF THE RECURRENCE OF RESECTABLE CANCER

◆7274

Gianni Bonadonna, M.D., Gabriele Tancini, M.D., Anna Rossi, M.D., and Marco Gasparini, M.D.
Istituto Nazionale Tumori, Milan 20133, Italy

INTRODUCTION

One of the most important new directions in clinical cancer research and management is the development of effective adjuvant systemic therapy to prevent recurrence in patients treated with local-regional modality. Although immunotherapy, endocrine therapy, and hyperthermia appear to show promise as adjuvants, chemotherapy with single or multiple agents is presently the most frequently used treatment for micrometastases (1). In the clinical area, the history of the combined modality approach began with the treatment of resectable Wilms' tumor in which the prolonged administration of adjuvant actinomycin D (Act D) markedly improved both the disease-free and the overall survival (2). For a number of years, the large majority of clinical oncologists failed to fully recognize that the successful strategic approach applied to Wilms' tumor was also worth testing in other neoplastic diseases at high risk of early relapse after optimal local-regional therapy. Thanks to information provided by animal tumor models, to improved knowledge on the natural history of several tumors, and to the advent of combination chemotherapy, new strategic approaches were recently proposed (3). Over the past five years innumerable programs have been designed with integration of systemic therapy, and in particular chemotherapy, into combined modality treatment for lymphomas and solid tumors. The aim of this review is to critically evaluate the early as well as most of the recent published results in the field of adjuvant chemotherapy for different forms of resectable cancer. To complete the section on pediatric tumors, the adjuvant therapy for Ewing's sarcoma is also reviewed.

0066-4219/78/0401-0149$01.00

PRINCIPLES OF ADJUVANT CHEMOTHERAPY

The philosophical base of the combined modality approach is the recognition that both radical surgery and radiotherapy have reached a plateau in their ability to cure solid tumors (3). The rationale for adjuvant chemotherapy, as derived from the experimental animal systems (4–7), can be summarized as follows: (*a*) The incidence of metastatic disease is directly related to tumor mass. (*b*) Surgical cure rates drop as tumor mass at surgery increases. (*c*) Grossly evident primary tumors are generally not curable by drug treatment. (*d*) Surgical adjuvant chemotherapy increases the long-term cure rates and significantly increases the life span of treatment failures. The study of population growth kinetics has revealed that one cancer cell must go through nearly 30 doublings (about 10^9 cells) for a single focus to be detectable by x ray or palpation. Therefore, by the time cancer is clinically detected, it is advanced and has had ample opportunity to establish distant micrometastases (8). Each micrometastasis contains from 10^8 down to 1 neoplastic cell, and the total body burden of tumor cells in micrometastases is the sum of all. In many cases this total body burden is in excess of 10^9 cells (6). Since to achieve cure it is necessary to reduce the total body burden to small numbers, perhaps even to less than one tumor cell, clinicians should be aware that the selective cell kill potential of the best currently available drugs or drug combinations is below the total body burden of tumor cells in over 90% of patients with overt metastatic tumors (6). Because the efficacy of chemotherapy is dose dependent and related to the body burden at the time of drug treatment, therapy has more chance to be effective when applied in full dosage and in the presence of a limited number of tumor cells (micrometastases). In metabolically homogeneous cell populations, the effective drug kill is best described by first-order reaction kinetics (4–6). Since the growth fraction is inversely related to the population size, micrometastases should be more vulnerable to the effects of cell-cycle specific agents (antimetabolites) (6).

In conclusion, the optimal kinetic conditions to achieve cure theoretically exist in most tumors when microscopic foci of disease are present after curative surgery and/or radiotherapy. The practical suggestions derived from the abovementioned studies are the following: (*a*) Systemic treatment is indicated in the subgroup of patients at high risk for treatment failure after local-regional modality. (*b*) Chemotherapy and/or immunotherapy should be started as soon as possible after surgery and/or radiotherapy. (*c*) The administration of adjuvant chemotherapy must be prolonged, since the long doubling time of solid tumors is due to a fraction of cells temporarily entering a nonproliferative state (G_O phase). In both experimental (4) and human (8) tumors it has been demonstrated that combinations of drugs are usually more effective than single agents and may eradicate 10 to 100 times as many tumor cells. Optimal scheduling is critical. In designing new study protocols, the investigators should also recall that drug response of the primary tumor may not reliably predict the sensitivity of micrometastases to cell-cycle specific drugs, and that optimal treatment scheduling for effective drugs against clinically advanced tumors may not predict optimal scheduling of the same drugs against micrometastases (5, 6). The above-mentioned concepts should increase the drugs available for consideration for adjuvant chemotherapy.

PEDIATRIC NEOPLASMS

Pediatric neoplasms, and in particular acute lymphoblastic leukemia, Wilms' tumor, and osteogenic sarcoma, have provided both in the past and more recently unparalleled models for developing new strategies in the treatment of human cancer. The aggressive nature of most pediatric tumors requires the use of chemotherapy even in the early stages of the disease. The frequently exquisite susceptibility to drug therapy has markedly contributed to defining complete response and to stressing the importance of duration of unmaintained remission. The achievements obtained with different forms of adjuvant chemotherapy are summarized in Table 1.

Wilms' Tumor

As stated in the introduction, adjuvant chemotherapy has substantially improved the prognosis of Wilms' tumor. In operable tumors surgery alone produced a two-year cure rate of 30% (9). Postoperative radiotherapy to the tumor bed increased the two-year survival rate to 47% (9). In a series of 1351 cases collected from the world literature and treated between 1940 and 1958, the two-year survival rate achieved with the above-mentioned modalities was considerably inferior, i.e. 20.9% and 25%, respectively (10). Farber (2) was the first investigator to report the usefulness of Act D administered intermittently every 3 mo for a period of 2 yr in clinically disease-free children after surgery and radiotherapy. His first detailed report showed a two-year disease-free survival rate of 89% (11). Subsequently, vincristine (VCR) was also proved to be an effective agent in the treatment of Wilms' tumor (12). Since then, the combination of these two drugs has been considered to be the best available adjuvant

Table 1 Disease-free survival (≥ 2 yr) in localized childhood tumors

Tumor	Survival with local modality (%)	Survival with adjuvant chemotherapy (%)	References
Wilms' tumor	25–45	60–90	2, 9–11, 13–15
Osteogenic sarcoma	10–30	>50	16, 19–24
Ewing's sarcoma	10–20	50–80	27–29, 33–35
Rhabdomyosarcoma	10–50	60–90	38–41,43–45
Medulloblastoma	25–40	65–85	50–54
Neuroblastoma	70–90	70–90	56, 57

therapy for resectable Wilms' tumor. In fact, in a controlled study with stages II and III of the disease, D'Angio et al (13) showed an improved actuarial two-year relapse-free (81%) and overall (86%) survival rate after combination chemotherapy compared to the results obtained with either Act D (57% and 67%) or VCR (55% and 72%) alone. However, the report of the Medical Research Council (14) showed excellent results in stages I, II, and III with VCR (relapse-free 87%, survival 93%) or Act D (62% and 80%) alone. Current studies are aimed at the improvement of cure rate through the addition of adriamycin (ADM) to the Act D and VCR regimens, and at the decrease of potential long-term morbidity from adjuvant treatment. In stage I, Act D provided the same results whether or not postoperative radiation was administered (13). In stages I and II, multiple courses vs a single five-day course of Act D (15) yielded the same overall survival rate, although prolonged chemotherapy produced a better disease-free survival rate (72%) than did short-term chemotherapy (43%).

Osteosarcoma

Osteosarcoma is a malignant tumor in which adjuvant chemotherapy has brought about significant changes in treatment attitudes and results. In clinically localized disease, the cure rate achieved with local treatment modality does not usually exceed 20%. In a review by Friedman & Carter (16), who have analyzed the results of amputation in 1337 patients, the five-year survival rate was 19.7% (range: 16–23%). Recently investigators at the Mayo Clinic observed a two-year survival rate of 45% (unpublished data). In 143 patients given high-dose radiation as primary treatment, the five-year survival rate was 12.5%, while in 396 cases treated with optimal radiotherapy followed by amputation, this finding was 19.8% (16). Most patients showed progressive clinical evidence of lung metastases during the first 12 mo with a median disease-free survival ranging from 6 to 10 mo. The new therapeutic approach was started when the high-dose methotrexate (HDMTX)–citrovorum factor rescue program (17) and ADM (18) were proven to be effective in 30–40% of patients with overt metastatic disease. The HDMTX regimen was progressively refined until it could be delivered on a weekly basis (19). In the Boston program, HDMTX is preceded by one dose of VCR, because this drug increases in vitro the uptake of tritiated MTX. In the absence of relapse, treatment is continued for 2 yr. The results of the first two studies were recently updated by Frei et al (19). In the historical control curve, 80% of patients have relapsed by 12 mo and the cure rate is between 15 and 20%. But in the group treated with HDMTX, the curve has plateaued at 50% and the minimum follow-up for this group of 12 patients is 3 yr. In the study with HDMTX plus ADM q. 4 wk, the curve at 3 yr is 67% with a minimum follow-up of 12 mo and a maximum follow-up of 36 mo. Cortes et al (20) treated 88 patients with ADM alone every 4–6 wk for a total of 6 courses (total 540 mg/m^2). In 47 patients adequately treated according to the protocol guidelines, 55% are expected to be disease-free at 60 mo (life table analysis). Protocol violations in terms of extent of amputation, dose of ADM, or both, occurred in 41 patients. In this group, 66% have relapsed. Thus in the whole treated group, 39% are expected to be disease-free at 60 mo. The efficacy of combination chemotherapy was first tested in a randomized study that employed the COMPADRI-I regimen [CTX, VCR, phenylalanine mustard (PAM), and ADM]. Of 43 patients, 56% are surviving with no evidence of disease for a period ranging from 12 to 61 mo. The COMPADRI-II regimen (same drugs + HDMTX) produced a 67% survival rate in 30 patients followed from 12 to 26 mo (21). The program was slightly modified in the COMPADRI-III regimen, but the results of this specific program are not yet available. Sutow (22) recently updated the results of the M.D. Anderson Hospital and reported an actuarial survival rate of 79% at 48 mo for 89 patients who received surgery plus adjuvant chemotherapy, compared to 20% of children who were treated only with amputation. Another multiple-drug regimen is ADM, cytoxan (CTX), and HDMTX for 10 mo (23). The median disease-free survival time of 20 patients is 15 mo, and 40% are alive with no evidence of disease at 20 mo from surgery as determined by the life table method. In osteosarcoma of adults, the Southwest Oncology Group (SOG) (24) adopted the CYVADIC regimen (CTX, VCR, ADM, and imidazole carboxamide or DTIC). Of 25 evaluable patients, 55% are disease-free survivors at 36 mo from surgery (actuarial analysis).

As the therapeutic effectiveness of chemotherapy has become established, its potential usefulness in the combined treatment of other aspects of osteosarcoma has been explored. One such investigation involves efforts to control the primary tumor in selected patients by local resection with prosthesis to avoid amputation and save the limb (25). Presurgical treatment includes intra-arterial ADM with or without HDMTX intravenously with or without radiotherapy. After segmental resection, further adjuvant chemotherapy is given. The initial results appear encouraging. However, they need to be confirmed on a larger number of patients followed for an adequate period of time. Another area of clinical activity is the continuing multimodal therapy (chemotherapy → resection of metastases → chemotherapy) to achieve response and even cure in patients who have developed pulmonary metastases (26).

Ewing's Sarcoma

Until a few years ago Ewing's sarcoma was almost always considered a uniformly fatal disease. The five-year survival rate of patients presenting with clinically localized disease and treated with surgery and/or radiotherapy varied from 4 to 19% (27, 28). Even in the presence of successful local control, the disease showed high tendency to recur in lungs and bones, and 75–85% of patients were dead within 2 yr of diagnosis (29). Pinkel (30) was the first to report the successful results of combined radiotherapy plus chemotherapy. Since 1970 numerous studies were undertaken with adjuvant chemotherapy, which utilized in various combinations those drugs (CTX, VCR, Act D, ADM) that were found effective in clinically advanced disease. Surgery was abandoned as local treatment (31) and new techniques of radiotherapy developed (32). Pomeroy & Johnson (33) combined local irradiation with progressively more intensive adjuvant chemotherapy. The last treatment utilized consisted of alternating high-dose pulses of ADM and CTX with VCR. Their results were most encouraging. Of 43 patients presenting with localized disease, 64% were alive at 2 yr and 52% at 5 yr. The most recent results of Memorial Hospital in New York (34), which used the T_2 study protocol (radiotherapy with Act D, ADM, VCR, and CTX) in 20 patients, showed an 80% two-year disease-free survival rate (median: 42 mo). At the Istituto Nazionale Tumori of Milan (35), radiation of the primary site is combined with intermittent cyclical courses of ADM, CTX, and VCR for a total of 18 mo. Of 19 consecutive patients, 14 are relapse-free for a period ranging from 12 to 40 mo, and the two-year actuarial disease-free survival rate is in excess of 50%. Other investigators have observed a two-year survival rate ranging from 50 to 70% by utilizing similar protocols. The advent of chemotherapy also improved the local control by radiotherapy (34). In an attempt to further improve the cure rate, the effectiveness of bilateral pulmonary irradiation is currently being evaluated in a controlled study by the Cancer Acute Leukemia Group B (ALGB).

The routine use of adjuvant radiotherapy to the central nervous system, as advocated by Pomeroy et al (33), is probably not justified. Current studies are trying to better define the optimal duration of adjuvant chemotherapy and to reduce the incidence of severe reactions (up to 30%) produced by combined treatment at the level of soft tissues surrounding the bone segment involved (34).

Rhabdomyosarcoma

Rhabdomyosarcoma is the commonest soft-tissue sarcoma in the pediatric age group. The pathology of this tumor (36, 37) was extensively studied in the 1950s, i.e. when anticancer drugs became available for clinical use. Among the four histologic types, the embryonal variety is that most often recognized in children. Adequate information about the natural history of radically resected rhabdomyosarcoma is scanty because of its rare incidence, the four histologic subgroups, and the various sites of primary lesion (head and neck area, retroperitoneum, pelvis, limbs). However, only 20–30% of patients are suitable for radical surgery at the time of diagnosis (38) and the cure rate by this modality alone ranges from 10 to 50%. In fact, Lawrence et al (39) reported 8 long-term disease-free survivors in a series of 32 patients with rhabdomyosarcoma of the head and neck area, and the Children Cancer Study Group (CCSG) reported 5 such survivors among 25 children radically operated on from 1960 to 1965 (40). The prognosis is only moderately improved by combining surgery with radiotherapy. Mahour et al (41) have obtained long-term survival in 10 of 49 patients utilizing this combined local modality. Pinkel in 1961 (42) was the first to propose adjuvant chemotherapy for clinically localized rhabdomyosarcoma. Since then the integrated multidisciplinary treatment has been gradually applied in numerous centers and has resulted in a substantial increase in the rate of survival. Kilman et al (43) observed a five-year survival rate of 71% in 31 patients treated with Act D and VCR. In those in whom the tumor was completely resectable, the survival rate was as high as 86%. In a controlled study carried out by CCSG on 32 children, the relapse rate after surgery alone was 55% compared to 17% of those treated with surgery + Act D + VCR (44). Ghavimi et al (45) utilized the T_2 protocol of Memorial Hospital (surgery, radiotherapy, and cyclical administration of Act D, ADM, VCR, CTX), and observed that all 20 patients with local-regional disease were alive at 4 to 42 mo from surgery.

The response to radiotherapy and chemotherapy now allows performance of a less mutilating surgery (46). In patients in whom radical surgery was not technically feasible, Heyn et al (47) reported a two-year survival rate of 70.8% in patients with microscopic residual disease and 43.2% in those with gross residual disease after treatment with irradiation + Act D + CTX + VCR. The cooperative groups (CCSG, ALGB, SOG) in the United States are presently combining their efforts in the Intergroup Rhabdomyosarcoma Study in an attempt to answer some critical questions about local-regional disease: (*a*) Is postoperative radiotherapy routinely indicated in surgically resectable tumors? (*b*) Is a three-drug regimen (Act D + CTX + VCR) really superior in the adjuvant situation to a two-drug combination (Act D + VCR)? (*c*) Should adjuvant chemotherapy be administered for one or two years?

Medulloblastoma

Surgical removal of the main tumor bulk facilitates the effectiveness of supplemental treatments with radiotherapy and chemotherapy. There is evidence that certain anticancer drugs (BCNU, CCNU, VCR, procarbazine, and intrathecal MTX) are effective in 40–100% of advanced cerebral tumors, including medulloblastoma (48, 49). Surgery followed by total neural axis radiation provides a five-year survival rate ranging from 25 to 40% (50, 51). To improve the cure rate, several programs were undertaken. The first utilized intrathecal radiogold. In 15 study patients a 75% three-year projected survival was observed by Fuller et al (52). Following high doses of external radiotherapy, Sheline (53) obtained a five-year disease-free survival in 7 of 8 children. In 1970 the Royal Marsden Hospital in London started a prospective study with surgery plus radiotherapy to cerebrospinal axis combined with CCNU, VCR, and intrathecal MTX (54). Of 14 treated patients, 65% were alive at 2 yr compared to 45% of 88 children treated from 1950 to 1970 with only surgery plus irradiation. Our own data compare favorably with the above-mentioned results. Of 13 children treated with surgery, irradiation (neural axis, 3000 rads; posterior fossa, up to 5000 rads), and combination chemotherapy (CCNU + VCR + intrathecal MTX) for two consecutive years, 12 children (92%) are disease-free from 1 to 5 yr (4 of 5 survivors were followed for more than 4 yr) (unpublished data). The CCSG is presently evaluating surgery plus radiotherapy vs the same modality followed by prednisone, CCNU, and VCR for a total of 8 cycles. The results are not yet available.

Neuroblastoma

Recent results (55) have shown that chemotherapy has not significantly influenced the mortality rate for this disease. Patients rarely present with stage I tumor (i.e. confined to the organ of origin), which is more often observed in children under one year of age. In the presence of resectable tumor, Koop (56) has reported a two-year survival rate of 84% using surgical therapy alone. No significant improvement was achieved by the addition of postoperative radiotherapy or chemotherapy or both. The only available study where the usefulness of adjuvant chemotherapy was evaluated on a controlled basis is that of CCSG (57). In 27 patients with stage I disease, CTX was administered after surgery for one year at the dose of 10 mg/kg through cycles of 7–10 days every 4 wk. No relapses were observed in either group (surgery alone, or surgery plus chemotherapy). Prolonged adjuvant chemotherapy with either single or multiple agents also failed to improve the relapse-free survival rate of children with regional or disseminated neuroblastoma when drugs were combined with radiotherapy (57, 58).

ADULT NEOPLASIAS

The most important forms of cancer in man, namely bronchogenic, mammary, and gastrointestinal carcinomas, occur in adults. In the presence of localized disease, surgery has been almost invariably the conventional form of primary treatment. Since the local-regional therapy is often followed by a progressive failure rate in distant sites, the contemporary strategic approach includes different types of systemic treatment. For other adult tumors the combined treatment modality has been consistently applied only in malignant melanoma and in glioblastoma multiforme, while new trials are in progress in testicular and ovarian carcinomas.

Bronchogenic Carcinoma

The catastrophic prognosis of bronchogenic carcinoma is well known and can be summarized as follows. The median survival of all patients from diagnosis to death remains less than 6 mo and the five-year survival rate is less than 10%. Half of the patients are inoperable at the time of diagnosis. Of the remaining half, 20% are found to be unresectable at thoracotomy. In resectable patients the overall five-year survival rate is approximately 25%. Therefore, even in this favorable group, 75% of the patients recur within 5 yr and the majority within the first 3 yr postoperatively (59, 60). Because of the short free interval and because the first relapse occurs almost invariably in extrathoracic sites, it is clear that the disease, and in particular small-cell carcinoma, is widely disseminated at the time of initial diagnosis. The results of the early adjuvant trials were recently reviewed (59–61). In short, CTX was the most commonly used single agent, followed by mechlorethamine (HN2) and vinblastine (VLB). Although a beneficial effect on survival was reported in a very few studies, the quality of study designs prevented rigorous analysis and firm conclusions. In fact, only the minority of trials have accounted for important variables affecting survival such as staging, age, sex, performance status, histological cell type, extent of surgical resection (lobectomy vs pneumonectomy), and the presence or absence of residual tumor after surgery. Furthermore, drug selection, and scheduling and intensity of treatment were often inadequate. Table 2 summarizes the results of the main adjuvant studies performed recently. The results of the early Swiss trial with CTX (70) were reevaluated in April 1977 with a median follow-up of 9 yr (62). In the treated group, 59 of 95 treated patients (77.6%) died from original carcinoma compared to 37 of 94 patients (55.2%) in the control group. Thus, the higher incidence of deaths in the CTX-treated group compared with the control group persisted unchanged since 1973 (71). The authors also observed that the survival curves of both groups did not converge with time, but remained parallel with the observed difference. The reasons for the unfavorable effects of CTX could be postulated as follows. First, in the two groups there was an uneven distribution of risk factors that were undetected by the analysis performed. Second, the drug not only failed to eliminate the residual disease, but with the dose-schedule utilized, it suppressed some unknown host defense mechanisms. In another study performed by Castberg (63), the short-term administration of CTX was also detrimental to the survival rate. These data were, however, not confirmed by a controlled study of the European Organization for Research on Treatment of Cancer (EORTC) group. (P. Levy, unpublished data). Table 2 also shows that most adjuvant trials with either single-agent or combination chemotherapy failed to favorably affect the survival rate. However, Katsuki et al (67) showed an improved survival rate primarily for the groups of patients with "undifferentiated" tumors and compared their results with a group of historical controls. Moreover, the results of the EORTC Lung Group (68) are still preliminary. In conclusion, the results so far obtained with different forms of chemotherapy for resectable non–small-cell carcinomas were generally either inconclusive or disappointing. New drugs, such as ADM, alone or in combination, chemotherapy plus immunostimulation, and, especially, well-designed

Table 2 Recent results in adjuvant chemotherapy for resectable non–small-cell bronchogenic carcinoma

Author (Reference)	Drug treatment[a]	Duration of treatment	Survival
Brunner (62)	CTX	2 yr	decreased
Castberg (63)	CTX	<1 mo	decreased
Levy[b]	CTX	2 yr	unchanged
Stott (64)	CTX or CLB	2 yr	unchanged
Castberg (63)	MTX	<1 mo	unchanged
Mountain (65)	CCNU	2 yr	unchanged
Higgins (66)	CTX + MTX	1.5 yr	unchanged
Higgins (66)	CCNU + HU	2 yr	unchanged
Katsuki (67)	Mito C + Chromo A_3	3 yr	increased
Israel (68)	CTX + MTX + CCNU (± BCG)	1 yr	unchanged
Karrer (69)	CTX + MTX + FU + VLB	2 yr	increased

[a] HU, Hydroxyurea; Mito C, Mitomycin; Chromo A_3, Chromomycin A_3; CLB, Chlorambucil
[b] Unpublished data of EORTC.

controlled trials are needed to substantially improve the treatment of resectable bronchogenic carcinoma.

Breast Carcinoma

Breast carcinoma has a notoriously high incidence, especially in Western countries. Despite innumerable studies and efforts to apply different forms of local-regional treatment, the prognosis has remained essentially unchanged over the past 30–40 years. The analysis of two controlled studies recently allowed accurate description of the natural history of operable breast cancer (72, 73). In patients with negative axillary lymph nodes (N−), prognosis is fairly good and, at 10 yr the relapse-free survival rate ranges from 72.1 to 75.9% and the overall survival rate from 64.9 to 81.9%. A recent study (74) showed that about half of the patients with tumor emboli to the intramammary lymphatics had early relapse in distant sites. In approximately 50% of women with operable breast cancer, the axillary nodes are histologically involved by metastases (N+). In this group prognosis is poor and is not influenced by postoperative radiotherapy. About half of the patients relapse within the first 3 yr from mastectomy and only 23.9–24.5% are disease-free survivors at 10 yr. When more than 3 lymph nodes are involved, the 10-year relapse-free survival rate is 13.8–16.4% and the overall survival rate is 13.4–25.6%. The corresponding figures for women with 1–3 positive nodes are 33.5–35.5% and 37.5–53.7%, respectively (72, 73).

During the past five years most research oncologists have become convinced that the best chance to improve the disease-free period and possibly the cure rate is first to identify those patients with a high probability of distant metastases at the time of curative surgery (e.g. N+) and then to apply a prolonged systemic treatment. At present there are few other aspects of cancer treatment that are more challenging to the skill and the emotions of investigators than the search for the control of micrometastases in breast cancer. The early trials were started in 1958 by B. Fisher and his colleagues of the National Surgical Adjuvant Breast Project (NSABP). Patients were randomized to receive thio-TEPA or placebo at the time of mastectomy and again on the first 2 postoperative days. The beneficial effects of thio-TEPA were limited to the subgroup of premenopausal women with more than 3 positive axillary nodes (75). After 10 years (72), 21% fewer patients in the thio-TEPA group had treatment failures and 21% more of them survived; thus the difference persisted in favor of women who received adjuvant chemotherapy. Other trials with single-agent chemotherapy [alkylating agents, fluorouracil (FU), mitomycin C] were subsequently carried out by other centers. However, in none was the study design such that the results were conclusive (61, 76). Based on the assumption that the cell kill by drugs occurred as zero-order reaction kinetics, chemotherapy was administered for short periods of time, and this was considered to represent the major cause of treatment failure. The new generation of adjuvant clinical trials began about five years ago. The strategy was based on the following concepts: (*a*) The N+ patients represent the subgroup at increased risk of relapse. (*b*) The control of local-regional recurrence with postoperative radiotherapy is not essential to improve the overall survival rate. (*c*) The cell kill by chemotherapy occurs by first-order reaction kinetics (4–6); therefore, drug treatment must be administered for a prolonged, yet undefined, period of time. (*d*) Combination chemotherapy in clinically disseminated breast cancer produces complete or partial remission in 50–75% of patients (77, 78) and responders show a better median survival (22–26 mo) than nonresponders (9–14 mo). (*e*) The technique of prospective randomized trials is essential to demonstrate the contribution of adjuvant chemotherapy. The early reports of the first studies carried out by the NSABP team (PAM for 2 yr vs placebo) and the Milan team [CTX, MTX, FU (CMF) for 12 mo vs control] showed that the relapse rate was statistically decreased (79, 80) at the expense of an acceptable toxicity. Both groups updated their results in February 1977 (81, 82). The NSABP findings at 24 mo continued to show a highly significant difference ($P = 0.009$) between the PAM and placebo groups. When subsets were analyzed, the percentage of reduction in treatment failure for postmenopausal women was no longer statistically significant, though the results were based on small numbers. The CMF results at 36 mo are summarized in Table 3. Also, in this study the difference between the total failure of treated and control women remains highly significant ($P < 0.0001$). The same applies for nodal status, premenopausal patients, and now also for the overall survival rate. However, in postmenopausal women the initial positive findings in terms of treatment failure are not statistically significant at 3 yr, although these data are based on a limited number of patients. Although the analysis failed to demonstrate that in menstruating patients the incidence of relapse was statistically influenced by CMF-induced amenorrhea (82), the more recent results of Rose & Davis (83) indicate that prolonged single-agent and combination chemotherapy suppress the ovarian function. Therefore, the beneficial effects of adjuvant chemotherapy in premenopausal women are probably achieved in part through endocrine manipulation. The importance of adjuvant endocrine treatment was recently stressed by Meakin et al (84). In a three-arm study on 756 evaluable patients, these authors found that after surgery and regional radiotherapy, ovarian radiation (2000 rads) plus prednisone (7.5 mg daily for up to 5 yr) resulted in a significant delay recurrence ($P = 0.01$) and

Table 3 Breast cancer with involved axillary lymph nodes. Three-year results after 12 cycles of adjuvant CMF (actuarial analysis)

Result	Controls (%) (179 patients)	CMF (%) (207 patients)	*P* value[a]
Total failure	46.2	29.2	<0.0001
Nodes, 1–3	37.8	23.2	<0.01
Nodes, >3	67.0	42.8	<0.001
Premenopause	48.0	16.0	<0.00001
Postmenopause	44.0	41.4	0.17
Overall survival	79.6	91.3	0.03

[a] Calculated on time distribution.

prolonged survival ($P = 0.02$) for the subgroup of premenopausal patients 45 yr or more with clinical stages I, II, and III. If only the patients with N+ were analyzed (75% in each group), the significance of the difference between no treatment and ovarian radiation plus prednisone decreased to $P = 0.03$ and $P = 0.06$, respectively.

Many other adjuvant studies are now in progress (1) to test different combinations, chemotherapy plus additive endocrine therapy (tamoxifen), and immunostimulation (BCG, *Corynebacterium parvum,* levamisole) alone or combined with cytotoxic drugs. However, due to either the limited number of patients accrued or to the short follow-up, it is premature to review their results. Nevertheless, two studies are worth a brief comment. In a controlled three-arm study (CMF, CMF + BCG, CMF + BCG + radiated tumor cell vaccine) involving 100 patients with an average follow-up of 14 mo, Sparks et al (85) observed a recurrence rate comparable to that initially reported with CMF (80). In an uncontrolled study, Budzar et al (86) utilized a combination of FU, ADM, CTX, and BCG in 75 patients with stages II and III. After 32 months of study, treatment failure occurred in 5.3% of treated patients as compared to 40% of 152 comparably matched control patients ($P = 0.016$). The advantage appeared statistically significant in all subgroups and toxicity was acceptable. Clearly, to demonstrate the advantage of immunostimulation, longer periods of time are required. Finally, in an uncontrolled study (R. G. Cooper, J. F. Holland, and O. Glidewell, in preparation), 73 patients with more than 3 axillary nodes received combination chemotherapy with CTX, MTX, FU, VCR, and prednisone for 9 mo. By life table analysis, 70% of the women were alive and disease-free at 7 yr. The curative impact of combination chemotherapy is therefore evident.

Colorectal Carcinoma

Large-bowel adenocarcinoma (of the colon and rectum) afflicts more patients than any other major malignant disease in the United States and in most countries of Europe. Radical surgical excision is the only universally accepted curative treatment for invasive colorectal carcinoma. Although the therapeutic results of surgical treatment remain difficult to interpret due to the lack of uniformity in the type of denominator used in published reports, survival rates have reached a plateau (87). The Dukes' classification remains the most practical method to predict the five-year survival rate after surgical modality (Class A; 61–81%; Class B; 25–64%; Class C; 6–28%). In advanced disease, FU, mitomycin C, and the nitrosourea derivatives were found to produce complete or partial remission in 10–20% of patients (87). The most effective combination regimen at present appears to be MeCCNU + FU + VCR. In a controlled study the response rate was 43.5% (88). In colorectal cancer the rationale for employing chemotherapy as an adjuvant to surgery has evolved significantly over the past 15 years. In the early trials the use of chemotherapeutic agents was limited to an intraoperative or immediate postoperative course in an attempt to sterilize the clonogenic cells liberated by surgical trauma. Only patients who had apparently curative resections were treated. In all these studies, therapy was designed to place the drug (thio-TEPA, mechlorethamine) as close as possible to the primary tumor site, i.e. intraperitoneally, intraluminally into the unresected bowel section, or via the portal vein. The use of alkylating agents failed to significantly increase the survival rate (61, 87, 89). The second generation of adjuvant trials was started with the use of fluorinated pyrimidines. In the first two controlled studies of the VA Surgical Adjuvant Cancer Chemotherapy Study Group, FU (90) and fluorodeoxyuridine (91) were administered during the postoperative period, i.e. within two months from surgery. No significant difference was observed in the survival rate between surgery alone and surgery followed by chemotherapy. Both trials utilized a large number of patients, but the unfavorable prognostic signs from the surgical-pathological point of view were poorly defined. In the third controlled study (90), the same research group has clearly outlined the criteria for starting adjuvant chemotherapy. These criteria, which were also adopted by most other groups, included: (*a*) positive regional lymph nodes; (*b*) serosal penetration or invasion of perirectal fat; (*c*) blood vessel invasion; (*d*) lymphatic invasion, and (*e*) cancer in other organs removed as part of the resection. In the study with prolonged intermittent FU therapy (five-day courses every 6 wk for 18 mo), no significant difference between the control group and the FU group was noticed in the survival rate of 522 evaluable patients (90). Also, Lawrence et al (92) failed to demonstrate significant benefit from an intensive adjuvant course of FU given over a one-year period first intraluminally, then intravenously, and subsequently by mouth. Another controlled study with FU administered intravenously for one year is that of Central Oncology Group (COG) (93). In 128 evaluable patients with Dukes' B and C classes, the survival rate was moderately more favorable than that of the control group, although the difference was not significant ($P = 0.09$). However, the treatment group showed a significantly better disease-free interval ($P = 0.045$). In a nonrandomized study, Li & Ross (94) recently reported that in 183 evaluable patients with Dukes' B and C classes, adjuvant FU improved the five-year survival rate compared to historical controls. Prolongation of postoperative disease-free interval and survival was claimed by Mavligit et al (95, 96) in 120 patients with Dukes' C class who received either BCG or BCG + oral FU for two years. In this study there was no difference between immunotherapy and chemoimmunotherapy, but both treatments were significantly better than historical, matched-control patients. The effectiveness of oral FU was nil in the randomized series of the Central Pennsylvania Oncology Group (97), although the reported findings were based on a limited number of patients. In conclusion, little, if any, benefit has yet been documented by adjuvant treatment with single agents given in a wide range of treatment schedules after potentially curative resection. Unfortunately, the only favorable reports have come from uncontrolled studies (94, 95).

The Gastrointestinal Tumor Study Group of the Division of Cancer Treatment of the National Cancer Institute (NCI), a contract-supported group of major cancer centers that perform combined modality studies on the major gastrointestinal adenocarcinomas, has recently initiated a series of trials in large-bowel cancer. Patients with operable rectal carcinoma (12 cm or less from the anal verge) and with Dukes' B_2 (extension to perirectal fat) and C classes are admitted to a four-arm study comparing (*a*) no additional treatment; (*b*) postoperative radiation with 4600 rads; (*c*) combination chemotherapy with FU + MeCCNU; (*d*) postoperative radiotherapy plus combination chemotherapy. In patients with colon cancer (more than 12 cm from the anal verge) immunotherapy will be substituted for the radiotherapy in an otherwise similar protocol design. Immunotherapy consists of a nonspecific immunostimulant, methanol extraction residue of BCG.

Gastric Carcinoma

Between 75 and 90% of all patients in whom stomach cancer is diagnosed are deemed suitable for laparotomy, and about 50–60% prove eligible for potentially curative resection. Surgical resection is at present the only therapy offering curative potential. The prognostic variables most associated with the five-year survival rate after radical surgery are lymph node involvement, gross type, and microscopic grading. The overall five-year survival rate ranges from 55 to 60% when the neoplasm is localized to the stomach. When the regional nodes are involved, the survival rate varies from 5 to 15%. Since these figures have changed little in the past 25 years, the need for an effective adjuvant treatment is great. In patients with advanced disease, FU and mitomycin C appear to be the most effective agents (response rate: 20–30%). Among other tested compounds, complete or partial remission was observed with mithramycin (20%), nitrosourea derivatives (18%), alkylating agents (17%), and ADM (15%). Most probably, chromomycin A_3 and olivomycin are moderately effective also (98). A series of controlled and uncontrolled studies (98, 99) has shown that combination chemotherapy is slightly superior to single-agent therapy. The most effective regimens (response rate: 40–55%) are those with BCNU or MeCCNU + FU; mitomycin C + FU + Ara-C; and FU + ADM + mitomycin C (98, 99).

The results of the early trials were recently well summarized by Carter (99) and Perloff (61). In general, most of the studies did not employ a prerandomization stratification and the chemotherapy regimens were usually of short duration and intensity. Essentially, no significant difference in overall survival or recurrence rate was obtained in seven trials using thio-TEPA, mitomycin C, fluorodeoxyuridine, CTX + chromomycin A_3, and nitromin. In a controlled study using FU, Blokhina et al (100) demonstrated a significant decrease in the incidence of recurrence and time to relapse for treated patients vs controls for the first 18 mo after surgery. The drug was administered at 15 mg/kg every other day for at least 3 courses during the first 3 mo after surgical resection. Fujimoto et al (101) reported the results of a controlled study using oral FU plus its furanyl analog FT-207. Treatment began on day 10 after the operation and continued for 24–36 mo without interruption. The comparative five-year survival rate between treated (57 cases) and control (78 cases) patients with serosal involvement was significant ($P < 0.005$). There was also a significant difference in the survival curves for the small number of patients with residual cancer infiltration on the cut edge of gastric resection. In the authors' opinion, this difference could be closely connected with high drug levels in the stomach after oral FU treatment.

The Gastrointestinal Tumor Study Group has two ongoing adjuvant protocols. The first study was designed to determine by controlled evaluation whether postoperative treatment with FU + MeCCNU for 2 yr or until relapse will substantially improve the cure rate, disease-free interval, total survival, and quality of life in patients treated by curative resection. The control group will receive no further therapy after surgical excision. In a second prospective randomized trial, the value of two therapies is compared, FU + MeCCNU vs a sequence of radiation therapy (5000 rads) + FU followed by the combination as adjuvant therapy for surgically incurable but localized gastric carcinoma.

Malignant Melanoma

The prognostic variables in malignant melanoma are primarily related to the histopathologic level of Clark and to the status of regional lymph nodes. Patients with regional node involvement have a five-year total survival rate of 15–25% and a disease-free survival rate of 6–8%, respectively, after surgical removal of all apparent disease. Those with clinically palpable, histologically positive nodes (N+), undergoing "therapeutic dissection," have a five-year survival rate of about 20%, whereas those with clinically nonpalpable, histologically positive nodes, subjected to "elective dissection," have a five-year survival rate of about 55% (102). Usually, treatment failure is not the result of uncontrolled local-regional disease, but instead is caused by clinically undetectable distant metastases. Thus, systemic adjuvant therapy is clearly warranted, especially in patients with clinically positive nodes.

A number of trials with adjuvant therapy were started during the past few years utilizing the agents [DTIC, nitrosourea derivatives, hydroxyurea (HU), Act D, BCG] that were found to produce objective response in overt metastatic disease. None of the study groups has accrued a sufficiently large number of patients or long-term follow-up evaluation adequate to provide definitive results on the effects of the various regimens employed. However, it can be anticipated that the trend is in favor of prolonged systemic adjuvant therapy. The studies with adjuvant BCG are well known (103–105) and their analysis is beyond the scope of this review. Hill et al (106) have tested DTIC vs control in 170 patients with positive regional nodes. The one-year disease-free survival rate was 67% for those who received DTIC and had thrombocytopenia, compared to 54% for controls and 35% for patients who, after DTIC, had no thrombocytopenia. Thus in this study the therapeutic advantage was associated with a certain degree of toxicity. Most of the available results include adjuvant treatment with chemoimmunotherapy. The overall survival rate after DTIC + BCG was superior to that after DTIC alone in the small N+ series of Ikonopisov (107). The Eastern Cooperative Oncology Group (ECOG) (T. Cunningham, unpublished data) randomized 100 patients with trunk melanoma between BCG weekly and DTIC

+ BCG. The one-year disease-free survival rate was similar in both groups. The lack of difference between BCG alone and DTIC + BCG was also noticed in the M. D. Anderson study (J. U. Gutterman, unpublished data) for patients with 5 or fewer positive nodes. In those with more than 5 positive nodes, there was no improvement with BCG. The WHO International Center for Melanoma has an ongoing prospective four-arm study. In 146 patients with N+ and previously untreated with surgery, the 18-month disease-free survival rate was inferior for control patients (32.5%) compared to treated groups (DTIC; 67.3%; BCG; 41.1%; DTIC + BCG; 59.9%). These unpublished results must still be considered very preliminary. Three studies provide some data on the effects of combination chemotherapy in N+ patients. McPherson et al (108), in 27 evaluable patients treated for 18 mo, found no statistical difference in the survival rate between DTIC + BCG and DTIC + BCNU + HU + BCG. However, the comparison of both treated groups with historical controls was significant at the $P = 0.01$ level. The French group (109), in a complex treatment protocol, utilized VLB, thio-TEPA, rufochromomycin, MTX, and procarbazine in patients with Clark's levels 3, 4, and 5. At 2 yr from surgery, 51% of patients treated only with surgery were relapse-free compared to 83% of those receiving systemic therapy. McMurtrey et al (110) carried out a pilot study with four treatment groups: 1. DTIC, MeCCNU, BCG; 2. DTIC, VCR, CTX, Act D, BCG; 3. DTIC, MeCCNU, BCG, melanoma viral oncolysates (MVO); 4. DTIC, VCR, CTX, Act D, MeCCNU, BCG, MVO. Treatment was administered from 6 to 24 mo. Eighty-two percent of patients remained clinically tumor-free in the chemoimmunotherapy groups 3 and 4, whereas, over 80% of untreated (after surgical removal of tumors) comparable historical controls and 60% of chemotherapy- and/or BCG-treated concurrent patients have succumbed to advancing tumors.

Malignant Gliomas

Glioblastoma multiforme (malignant astrocytoma, astrocytoma grades III and IV) is the most common glioma, accounting for about 55% of all gliomas. Surgery (wide resection of tumor or decompression of brain) represents the treatment of choice. High-energy radiation (4500–6000 rads in 5–8 wk) is commonly employed in the postoperative period, or instead of surgery when resection is not technically feasible. The median postoperative survival among patients with glioblastoma multiforme is 4.5 mo and only 10% of patients are still alive at 18 mo.

The era of neurochemotherapy started with the introduction of liposoluble agents such as the nitrosourea derivatives. Adjuvant chemotherapy of brain tumors is directed to macro- as well as to microdeposits of malignant cells left after maximal surgical decompression. The first important controlled study was that of the Brain Tumor Study Group (BTSG) in the United States (111). After surgical resection, 223 patients were randomized into 4 groups. The best median survival was that of the group treated with radiotherapy + BCNU (40 wk), followed by those given radiotherapy (37 wk), BCNU (25 wk), and no further therapy (17 wk). In a similar study carried out in Milan on 83 patients (C. L. Solero et al, in preparation), the median survival from surgery was longer in the group given radiotherapy + CCNU (78 wk) compared to radiotherapy + BCNU or radiotherapy alone (52 wk). The difference was significant ($P = 0.05$). Also, Band et al (112) found a slight but not significant increase in the median survival with radiotherapy + CCNU (36 wk) compared to either radiotherapy (25 wk) or CCNU (33 wk). Results favoring radiotherapy + CCNU over radiotherapy alone were also observed by other investigators (113, 114). On the contrary, M. S. Mahaley (unpublished data) and Walker (115) reported a superiority of BCNU over MeCCNU. In particular, the most recent trial of the BTSG on 332 patients showed a median survival of 51 wk for radiotherapy + BCNU compared to 31 wk for either radiotherapy + MeCCNU or MeCCNU alone. The group given only radiotherapy after surgery showed a median survival of 35 wk. Other studies include adjuvant combination chemotherapy (116) and chemo-immunotherapy with BCG (117). The results are not yet fully available.

Testicular Carcinomas

Testicular neoplasms are rare. While extensive radiotherapy represents the treatment of choice following orchiectomy in pure seminoma, the treatment of the other germinal tumors remains a topic of controversy because of histologic classification, extent of staging procedures, and best combined therapeutic approach. There are no doubts that the spread of the disease to retroperitoneal nodes markedly decreases the five-year survival rate. In patients subjected to combined orchiectomy and accurate retroperitoneal lymph node dissection, the survival rate is 75–87% in N− patients compared to 50–60% for those with N+. In addition, radiotherapy can provide comparable results in the presence of very small (< 2 cm) or microscopic nodal metastases. Combination chemotherapy for overt clinical metastatic disease has dramatically improved during the past five years. New combinations were designed and the high incidence of complete remission (40–75%) was often translated into a long-term survival and even cure (118, 119). A greater response is therefore expected in the presence of micrometastases. It is now accepted that adjuvant chemotherapy is required beyond the surgical approach once the retroperitoneal nodes are involved and operable. Few reports on radical surgery (N+ and N− patients) followed by systemic therapy are presently available.

The results of Ansfield & Ramirez (120) can be summarized as follows. The triple combination with chlorambucil, MTX, and Act D was given to 32 patients for 1–2½ yr. Of these patients, 4 had choriocarcinoma and all were alive and disease-free after 4–12 yr. This is important, as prior to adjuvant chemotherapy this disease was almost invariably fatal. Of 17 patients with embryonal carcinoma, 12 had lymphadenectomy, which was positive in 7. Of the 17, 11 are tumor-free for more than 3 yr. Of 11 with teratocarcinoma, 4 had positive nodes and 10 are alive with no evidence of disease for more than 3 yr. Pommatau et al (121) administered for at least one year a similar adjuvant protocol where CTX was substituted for chlorambucil. Of 17 patients with N+ after lymphadenectomy (10 also had postoperative radiotherapy), 9 are disease free from a minimum of 2.5 to a maximum of 11 yr. Postoperative chemotherapy alone was administered to 11 N− patients and 10 are alive without recurrence from 3 to 11 yr. Utilizing a complex treatment protocol with ADM, VCR, bleomycin (BLM), and CTX followed by VCR, BLM, MTX, and then by VCR + MTX or Act D, Amiel et al (122) treated a total of 6 patients immediately after lymphadenectomy. The 2 patients with choriocarcinoma remain in complete remission for 1 and 3 yr, respectively, while 3 of 4 with embryonal carcinoma are alive and disease free at 20, 37, and 90 mo from surgery, respectively. Unfortunately, the nodal status, whether N− or N+, is not known. Samuels et al (119) treated 10 patients with resectable retroperitoneal nodes with adjuvant VLB + BLM. Of these, 9 are living with no evidence of disease and have a survival rate ranging from 1.8 to 3.3 yr; the tenth patient died because of radiation injury to the bowel.

As in osteogenic sarcoma, the potential effectiveness of combined treatment has also been explored in the management of more advanced stages. Patients are first brought into either complete or good partial remission by intensive chemotherapy. Then surgery is carried out to remove residual nodal and/or extranodal tumors. The initial reports (123, 124) indicate that this is a promising new direction in therapy. Because of the paucity of testicular tumors, cooperative groups, such as the Testicular Tumor Study Group organized by NCI, recently began controlled trials to answer specific questions related to optimal treatment modalities for each stage.

Ovarian Carcinoma

Carcinoma of the ovary is the leading cause of gynecological cancer deaths. The main reason is that approximately 70% of patients initially present with advanced disease outside the confines of the true pelvis (FIGO stages III and IV). In addition, at least one third of women with localized disease (FIGO stages I and II) and treated with surgery and/or radiation develop recurrent disease. Despite advances made in the staging and in the medical treatment of ovarian cancer (125), very few studies have been undertaken with postoperative adjuvant chemotherapy. In the series of Griffiths et al (126), patients with stages II and III of the disease first had an "optimal operation," defined as the excision of more than 50% of the tumor bulk leaving no masses large enough to palpate postoperatively. A total of 25 women received either chlorambucil or CTX concurrently with postoperative radiation. Since leukopenia frequently necessitated interruption of therapy, the drug was continued for 1–4 wk after radiation in 8 patients. The group who received radiation therapy plus chemotherapy showed an improved free interval and an increased median survival time compared to the group treated only with postoperative radiation. In another nonrandomized study, women with stages I and IIa received postoperative chlorambucil. Only 1 of 17 patients with stage I and II of 6 patients with stage IIa developed recurrent disease (127). However, this series is too small to draw conclusions.

Single-agent chemotherapy with PAM was given postoperatively in two prospective randomized trials. Smith et al (128), after initial surgery to remove as much tumor as possible (residual masses < 2 cm and no ascites), randomized 149 patients between postoperative radiation to the entire abdomen and PAM for 12 monthly cycles. No stratification for stage was made. At the end of adjuvant chemotherapy all patients had a "second look" operation compared to only 3 of those given radiotherapy. The comparative two-year relapse-free survival rate between women treated with radiotherapy and those treated with PAM was as follows: stage I, 85% vs 90%; stage II, 55% vs 58%; stage III, 53% vs 35%. When all stages were combined, 63% of women in both treatment groups were still in complete remission. However, at 5 yr the disease-free status was in favor of chemotherapy (63%) rather than radiotherapy (45%). The overall survival rate curves for the two groups were similar. The authors concluded that neither adjuvant treatment method seemed to offer a clear-cut superiority. However, radiotherapy was more toxic, time consuming, and expensive than cyclical PAM chemotherapy. The second controlled study is that of the Gynecologic Oncology Group (129). Patients with stage Ia (growth limited to one ovary; no ascites) and Ib (growth limited to two ovaries; no ascites), within 4 wk from laparotomy with total hysterectomy and bilateral salpingo-oophorectomy, were randomized to no further treatment (regimen 1), radiation with 5000 rads (regimen 2), or PAM for 18 monthly cycles (regimen 3). The updated results (130) revealed no statistical difference in the relapse rate among the three treatments groups (regimen 1, 5 of 39; regimen 2, 1 of 35; regimen 3, 3 of 38).

Miscellaneous Tumors

In a number of malignant tumors the combined local-systemic treatment approach has as yet received little attention. Although new adjuvant treatment techniques have been developed recently, few reports are presently available on this specific subject.

In soft-tissue sarcomas of adults even when local control is achieved, over 50% of the patients eventually develop distant metastases. Morton et al (131, 132) treated a small group of patients with soft-tissue sarcomas of the extremity with either surgery followed by ADM + HDMTX, or with preoperative intraarterial ADM + irradiation followed by surgery and further ADM + HDMTX (with or without immunotherapy) for 4–28 mo. The early results appear encouraging and, by use of preoperative treatment, limb salvage was accomplished in 94% of patients. After surgery, Maurer (133) randomized patients with stage I disease to VCR + Act D + CTX (21 cases), or to radiotherapy plus the same chemotherapy (10 cases). Of patients in both groups, 90% remained disease free from 1 to 119 wk.

In squamous carcinomas of the head and neck area, multimodality treatment has been utilized for a number of years (134). However, because of the tendency of this type of cancer to recur in the local-regional area and because a consistent number of patients present with primary inoperable disease, chemotherapy with MTX with or without leucovorin, BLM, or ADM was practically always used intraarterially or intravenously before surgery and/or radiotherapy in the attempt to reduce the bulky neoplasm. Recently, Taylor et al (135) randomized 30 patients to receive, after surgery and radiotherapy, either MTX + leucovorin (2-wk courses repeated every 3 mo for 1 yr) or the same chemotherapy plus an autochthonous tumor cell vaccine and BCG. Of the 12 patients achieving complete remission in the chemotherapy group, 8 continued with no evidence of disease up to about 20 mo. This finding was observed in 10 of 12 patients receiving chemoimmunotherapy.

Anaplastic carcinoma of the thyroid carries a very severe prognosis. Adjuvant chemotherapy appears indicated especially because the disease is definitely responsive to ADM. The only available report on a multimodality program for high-risk thyroid carcinoma is that of Durie et al (136). After sub–total thyroidectomy, a complex induction treatment was started with ^{131}I (150 mCi), ADM, PAM, VCR, and BLM. Ten courses of the combination therapy were given. Maintenance treatment included PAM + BCG for 1 yr. All 4 patients with anaplastic carcinoma showed evidence of response and their median survival was 20 mo. Similar results were observed in follicular carcinoma with vascular invasion.

In endometrial cancer, the only reported adjuvant study (137) used medroxyprogesterone vs placebo in stage I of the disease (growth limited to the corpus). The hormone was given before hysterectomy and then weekly for 14 wk. At 4 yr, no improvement in the disease-free survival was observed in the group treated with medroxyprogesterone (87%) compared to that receiving placebo (92%).

No results on postoperative adjuvant chemotherapy are presently available for the following tumors: chondrosarcoma, carcinomas of the esophagus, pancreas, kidney, urinary bladder, and uterine cervix.

CONCLUSION

During the past decade the medical treatment for cancer has markedly improved its potential in many neoplastic diseases through the discovery of new drugs, the development of cyclical combination chemotherapy, the application of supportive measures, and, more recently, the advent of combined local-systemic modalities. Once, drug therapy was relegated to palliative treatment of overt disseminated disease. Now, cure is the goal of cancer chemotherapy and chemoimmunotherapy. This is being obtained in an increasing number of patients either in the presence of advanced tumors achieving complete remission after intensive combination chemotherapy (e.g. hematologic malignancies, Wilms' tumor, embryonal rhabdomyosarcoma, testicular carcinoma), or when drugs are administered under optimal strategic conditions, i.e. once surgery and/or radiotherapy have eradicated the primary local-regional disease. Much remains to be done to develop more effective medical treatment, determine optimal drug sequence, and modulate systemic therapy in selected subgroups of patients. However, present achievements in certain neoplastic diseases of adults and children have already been made in two major areas. The first is the demonstration that the strategic principles of cell kill by drugs, as developed in animal model systems, are being translated into clinical evidence by prospective trials. The second, and even more important, result is that clinicians are gradually becoming convinced that recent observations made on the natural history of most neoplastic diseases, following an optimal local modality, should be interpreted with a different therapeutic attitude. Therefore, new strategies must be developed and tested in high-risk patients, even if systemic treatment should be associated with acute or delayed, predictable or as yet undefined morbidity. The next decade will probably represent the most critical and significant period in the history of treatment of resectable cancer. Despite the undoubted success in treating certain forms of cancer, few treatments have yet become conventional and most remain experimental. For this reason, the scientific hypotheses and the biological questions that involve the new multimodality approach for the treatment of cancer must be initially tested through simple, well-designed, prospective controlled trials. As correctly stated by P. P. Carbone, "Research results are hard to predict and more often than not they may be contrary to what is anticipated. Only time and clinical trials will give us the answer to all our questions. The answers will come as well as new questions to be raised" (77). In our opinion, this appears to be the most appropriate working message for research oncologists who are willing to accept new ideas and new findings.

Literature Cited

1. Salmon, S. S., Jones, S. E., eds. 1977. *Adjuvant Therapy of Cancer.* Amsterdam: North-Holland. 646 pp.
2. Farber, S., D'Angio, G. J., Evans, A., Mitus, A. 1960. Clinical studies with actinomycin D with special reference to Wilms' tumor in children. *Ann. NY Acad. Sci.* 89:421–24
3. Carter, S. K., Soper, W. T. 1974. Integration of chemotherapy into combined modality treatment of solid tumors. I. The overall strategy. *Cancer Treatment Rev.* 1:1–13
4. Skipper, H. E., Schabel, F. M. Jr. 1977. In *Cancer Medicine,* ed. J. F. Holland, E. Frei III. Philadelphia: Lea & Febiger. 2nd ed. In press
5. Schabel, F. M. Jr. 1975. Concepts for systemic treatment of micrometastases. *Cancer* 35:15–24
6. Schabel, F. M. Jr. 1977. Rationale for adjuvant chemotherapy. *Cancer* 39:Suppl., pp. 2875–82
7. Martin, D. S., Fugmann, R., Stolfi, R., Hayworth, P. 1974. Rationale of combined modality therapy of cancer. *Cancer Chemother. Rep.* 4:13–24
8. De Vita, V. T. Jr., Young, R. C., Canellos, G. P. 1975. Combination versus single agent chemotherapy: a review of the basis for selection of drug treatment of cancer. *Cancer* 35:98–110
9. Gross, R. E., Neuhauser, E. B. D. 1950. Treatment of mixed tumors of the kidney in childhood. *Pediatrics* 6:843–52
10. Klapproth, H. J. 1959. Wilms' tumor: a report of 45 cases and analysis of 1,351 cases reported in the world literature from 1940 to 1958. *J. Urol.* 81:633–48
11. Farber, S. 1966. Chemotherapy in the treatment of leukemia and Wilms' tumor. *J. Am. Med. Assoc.* 198:826–36
12. Sullivan, M. P., Sutow, W. W., Cangir, A., Taylor, G. 1967. Vincristine sulfate in the management of Wilms' tumor. *J. Am. Med. Assoc.* 202:381–84
13. D'Angio, G. J., Evans, A. E., Breslow, N., Beckwith, B., Bishop, H., Faigl, P., Goodwin, W., Leape, L. L., Sinks, L. F., Sutow, W. W., Tefft, M., Wolff, J. 1976. The treatment of Wilms' tumor. *Cancer* 38:633–46
14. Jenkin, R. D. T. 1976. The treat-

ment of Wilms' tumor. *Pediatr. Clin. North Am.* 23:147–60
15. Wolff, J. A., D'Angio, G., Hartmann, J., Krivit, W., Newton, W. A. Jr. 1974. Long-term evaluation of single versus multiple courses of actinomycin D therapy of Wilms' tumor. *N. Engl. J. Med.* 290:84–86
16. Friedman, M. A., Carter, S. K. 1972. The therapy of osteogenic sarcoma: current status and thoughts for the future. *J. Surg. Oncol.* 4:482–510
17. Jaffe, N. 1972. Recent advances in the chemotherapy of metastatic osteogenic sarcoma. *Cancer* 30:1627–31
18. Cortes, E. P., Holland, J. F., Wang, J. J., Sinks, L. F. 1972. Doxorubicin in disseminated osteosarcoma. *J. Am. Med. Assoc.* 221:1132–38
19. Frei, E. III, Jaffe, N., Skipper, H. E., Gero, M. G. 1977. See Ref. 1, pp. 49–64
20. Cortes, E. P., Holland, J. F., Glidewell, O. 1977. See Ref. 1, pp. 407–23
21. Sutow, W. W., Gehan, E. A., Vietti, T. J., Frias, A. E., Dyment, P. G. 1976. Multidrug chemotherapy in primary treatment of osteosarcoma. *J. Bone Jt. Surg.* 58(A):629–33
22. Sutow, W. W. 1977. See Ref. 1, pp. 345–56
23. Pratt, C., Shanks, E., Hustu, O., Rivera, G., Smith, J., Kumar, A. P. M. 1977. Adjuvant multiple drug chemotherapy for osteosarcoma of the extremity. *Cancer* 39:51–57
24. Murphy, W. K., Benjamin, R. S., Eyre, H. J., Thigpen, T., Whitecar, J. P. Jr., Uribe-Botero, G., Baker, L. H., Gehan, E. A., Gottlieb, J. A. 1977. See Ref. 1, pp. 399–406
25. Rosen, G., Murphy, M. L., Huvos, A. G., Gutierrez, M., Marcove, R. C. 1976. Chemotherapy, en bloc resection, and prosthetic bone replacement in the treatment of osteogenic sarcoma. *Cancer* 37:1–11
26. Jaffe, N., Frei, E. III, Traggis, D., Watts, H. 1977. Weekly high-dose methotrexate-citrovorum factor in osteogenic sarcoma. *Cancer* 39:45–50
27. Falk, S., Alpert, M. 1967. Five year survival of patients with Ewing's sarcoma. *Surg. Gynecol. Obstet.* 124:319–24
28. Dahlin, D. C., Coventry, M. B., Scanlon, P. W. 1961. Ewing's sarcoma. A critical analysis of 165 cases. *J. Bone Jt. Surg.* 43(A):185–92
29. Phillips, T. L., Sheline, G. E. 1969. Radiation therapy of malignant bone tumors. *Radiology* 92:1537–45
30. Pinkel, D. 1962. Cyclophosphamide in children with cancer. *Cancer* 15:42–49
31. Boyer, C. W. Jr., Brickner, T. J. Jr., Perry, R. H. 1967. Ewing's sarcoma: case against surgery. *Cancer* 20:1602–6
32. Suit, H. D., Martin, R. G., Sutow, W. W. 1973. In *Medical Pediatric Oncology*, ed. W. W. Sutow, T. J. Vietti, D. J. Fernbach, pp. 473–96. St. Louis: Mosby. 602 pp.
33. Pomeroy, T. C., Johnson, R. E. 1975. Combined modality therapy for Ewing's sarcoma. *Cancer* 35:36–47
34. Tefft, M., Chabora, B. M., Rosen, G. 1977. Radiation in bone sarcomas. *Cancer* 39:806–16
35. Fossati-Bellani, F., Barni, S., Gasparini, M., Lombardi, F., Lattuada, A., Bonadonna, G. 1977. See Ref. 1, pp. 373–80
36. Stout, A. B. 1946. Rhabdomyosarcoma of the skeletal muscles. *Ann. Surg.* 123:447–53
37. Horn, R. C., Enterline, H. T. 1958. Rhabdomyosarcoma: a clinico-pathological study and classification of 39 cases. *Cancer* 11:181–99
38. Heyn, R. M. 1975. The role of chemotherapy in the management of soft tissue sarcomas. *Cancer* 35:921–24
39. Lawrence, W. Jr., Jegge, G., Foote, F. W. Jr. 1964. Embryonal rhabdomyosarcoma. A clinico-pathological study. *Cancer* 17:361–76
40. Johnson, D. G. 1975. Trends in surgery for childhood rhabdomyosarcoma. *Cancer* 35:916–20
41. Mahour, G. H., Soule, E. H., Mills, S. D., Lynn, H. B. 1967. Rhabdomyosarcoma in infants and children: a clinico-pathologic study of 75 cases. *J. Pediatr. Surg.* 2:402–9
42. Pinkel, D., Pickren, J. 1961. Rhabdomyosarcoma in children. *J. Am. Med. Assoc.* 175:293–98
43. Kilman, J. W., Clatworthy, H. W. Jr., Newton, W. A. Jr., Grosfeld, J. L. 1973. Reasonable surgery for

rhabdomyosarcoma: a study of 67 cases. *Ann. Surg.* 178:346–51

44. Heyn, R., Holland, R. 1970. Treatment of rhabdomyosarcoma in children. *Proc. Am. Assoc. Cancer Res.* 11:36 (Abstr.)
45. Ghavimi, F., Exelby, P. R., D'Angio, G. J., Cham, W., Lieberman, P. H., Tan, C., Miké, V., Murphy, M. L. 1975. Multidisciplinary treatment of embryonal rhabdomyosarcoma in children. *Cancer* 35:677–86
46. Tefft, M., Fernandez, C. H., Moon, T. E. 1977. Rhabdomyosarcoma: response with chemotherapy prior to radiation in patients with gross residual disease. *Cancer* 39:665–70
47. Heyn, R., Holland, R., Joo, P., Johnson, D., Newton, W. Jr., Tefft, M., Breslow, N., Hammond, D. 1977. Treatment of rhabdomyosarcoma in children with surgery, radiotherapy and chemotherapy. *Med. Pediatr. Oncol.* 3:21–32
48. Broder, L. E., Rale, D. B. 1972. Chemotherapy of brain tumors. *Prog. Exp. Tumor Res.* 17:373–99
49. Shapiro, W. R. 1975. Chemotherapy of primary malignant brain tumors in children. *Cancer* 35:965–72
50. Bloom, H. J. G., Wallace, E. N. K., Henk, J. M. 1969. Treatment and prognosis of medulloblastoma in children: study of 82 verified cases. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* 105:43–62
51. Jenkin, R. D. T. 1969. Radiation therapy of medulloblastoma in childhood. *Can. Med. Assoc. J.* 100:51–53
52. Fuller, A., Rodoff, E., Deck, M., Galicich, J., Ghavimi, F., Levitt, S., Smith, C. E., D'Angio, G. J. 1974. Recent experience with intrathecal radiogold for medulloblastoma and ependymoblastoma: a progress report. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* 122:75–79
53. Sheline, G. E. 1975. Radiation therapy of tumors of the central nervous system in childhood. *Cancer* 35:957–64
54. Bloom, H. J. G. 1975. Combined modality therapy for intracranial tumors. *Cancer* 35:111–20
55. Gasparini, M., Fossati-Bellani, F., Musumeci, R., Bonadonna, G. 1974. Response and survival of patients with metastatic neuroblastoma after combination chemotherapy with adriamycin (NSC 123127), cyclophosphamide (NSC 26271), and vincristine (NSC 67574). *Cancer Chemother. Rep.* 56:499–503
56. Koop, C. E. 1968. Neuroblastoma: two year survival and treatment correlations. *J. Pediatr. Surg.* 3:178–79
57. Evans, A. E., Albo, V., D'Angio, G. J., Finklestein, J. Z., Leiken, S., Santulli, T., Weiner, J., Hammond, G. D. 1976. Cyclophosphamide treatment of patients with localized and regional neuroblastoma. *Cancer* 38:655–60
58. Green, A. A., Hustu, H. O., Palmer, R., Pinkel, D. 1976. Total-body sequential segmental irradiation and combination chemotherapy for children with disseminated neuroblastoma. *Cancer* 38:2250–57
59. Legha, S. S., Muggia, F. M., Carter, S. K. 1977. Adjuvant chemotherapy in lung cancer: review and prospects. *Cancer* 39:1415–24
60. Livingston, R. B. 1977. See Ref. 1, pp. 191–205
61. Perloff, M., Holland, J. F. 1977. Surgical adjuvant chemotherapy. *Ann. Rev. Med.* 28:475–88
62. Brunner, K. W., Marthaler, T. H., Müller, W., Alberto, P. 1977. In *Second National Cancer Institute Conference on Lung Cancer Treatment*, ed. F. M. Muggia, M. Rozencweig. New York: Raven. In press
63. Castberg, T., Kruse-Blinkenberg, H. O. 1976. Cancer chemotherapy as a supplement to operation in bronchogenic carcinoma. *Dan. Med. Bull.* 23:254–55
64. Stott, H., Stephens, R. J., Fox, W., Roy, D. C. 1976. 5-year follow-up of cytotoxic chemotherapy as an adjuvant to surgery in carcinoma of the bronchus. *Br. J. Cancer* 34:167–73
65. Mountain, C. F. 1977. See Ref. 62. In press
66. Higgins, G. A. 1977. See Ref. 62. In press
67. Katsuki, H., Shimada, K., Koyama, A., Okita, M., Yamaguchi, Y., Okamoto, T. 1975. Long-term intermittent adjuvant chemotherapy for primary, resected lung cancer. *J. Thorac. Cardiovasc. Surg.* 70:590–99
68. Israel, L., Bonadonna, G., Sylvester, R., and EORTC Members. 1977. See Ref. 62. In press

69. Karrer, K., Pridun, M., Denck, H. 1977. See Ref. 62. In press
70. Brunner, K. W., Marthaler, T. H., Müller, W. 1971. Unfavourable effects of long-term adjuvant chemotherapy with Endoxan in radically operated bronchogenic carcinoma. *Eur. J. Cancer* 7:285–94
71. Brunner, K. W., Marthaler, T. H., Müller, W. 1973. Effects of long-term adjuvant chemotherapy with cyclophosphamide (NSC 26271) for radically resected bronchogenic carcinoma. *Cancer Chemother. Rep.*, Pt. 3, 4:125–32
72. Fisher, B., Slack, N., Katrych, D., Wolmark, N. 1975. Ten year follow-up results of patients with carcinoma of the breast in a co-operative clinical trial evaluating surgical adjuvant chemotherapy. *Surg. Gynecol. Obstet.* 140:528–34
73. Valagussa, P., Bonadonna, G., Veronesi, U. 1977. Importance of axillary nodes in selecting breast cancer patients for adjuvant therapy. *Proc. Am. Assoc. Cancer Res.* 18:282 (Abstr.)
74. Nime, F. A., Rosen, P. P., Thaler, H., Ashikary, R., Urban, J. A. 1977. Prognostic significance of tumor emboli in intramammary lymphatics in patients with mammary carcinoma. *Am. J. Surg. Pathol.*, Vol. 1. In press
75. Fisher, B., Ravdin, R. G., Ausmann, R. K., Slack, N. H., Moore, G. E., Noer, R. J. 1968. Surgical adjuvant chemotherapy in cancer of the breast. *Ann. Surg.* 168:337–56
76. Tormey, D. C. 1975. Combined chemotherapy and surgery in breast cancer: a review. *Cancer* 36:881–92
77. Carbone, P. P. 1976. Systemic treatment of breast cancer: past, present and future. *Int. J. Radiat. Oncol. Biol. Phys.* 1:759–67
78. Haskell, C. M., Sparks, F. C., Graze, P. R., Korenmann, S. G. 1977. Systemic therapy for metastatic breast cancer. *Ann. Int. Med.* 86:68–80
79. Fisher, B., Carbone, P., Economou, S. G., Frelick, R., Glass, A., Lerner, H., Redmond, C., Zelen, M., Band, P., Katryck, L. D., Wolmark, N., Fisher, E., and other cooperating investigators. 1975. L-Phenylalanine mustard (L-PAM) in the management of primary breast cancer. A report of early findings. *N. Engl. J. Med.* 292:117–22
80. Bonadonna, G., Brusamolino, E., Valagussa, P., Rossi, A., Brugnatelli, L., Brambilla, C., De Lena, M., Tancini, G., Bajetta, E., Musumeci, R., Veronesi, U. 1976. Combination chemotherapy as an adjuvant treatment in operable breast cancer. *N. Engl. J. Med.* 294:405–10
81. Fisher, B., Redmond, C., and participating NSABP investigators. 1977. See Ref. 1, pp. 67–81
82. Bonadonna, G., Rossi, A., Valagussa, P., Banfi, A., Veronesi, U. 1977. See Ref. 1, pp. 83–94
83. Rose, D. P., Davis, T. E. 1977. Ovarian function in patients receiving adjuvant chemotherapy for breast cancer. *Lancet* 1:1174–76
84. Meakin, J. W., Allt, W. E. C., Beale, F. A., Brown, T. C., Bush, R. S., Clark, R. M., Fitzpatrick, P. J., Hawkins, N. V., Jenkin, R. D. T., Pringle, J. F., Rider, W. D., Hayward, J. L., Bulbrook, R. D. 1977. See Ref. 1, pp. 95–99
85. Sparks, F. C., Meyerowitz, B. E., Ramming, K. P., Wolk, R. W., Goldsmith, M. H., Lemkin, R. S. R., Spears, I. K., Morton, D. L. 1977. See Ref. 1, pp. 109–14
86. Budzar, A. V., Blumenschein, G. R., Guttermann, J. U., Tashima, C. K., Hortobagyi, G. N., Wheeler, W., Gehan, E., Freireich, E. J., Hersh, E. See Ref. 1, pp. 139–46
87. Carter, S. K. 1976. Large bowel cancer. The current status of treatment. *J. Natl. Cancer Inst.* 56:3–10
88. Moertel, C. G., Schutt, A. J., Hahn, R. G., Reitemeier, R. J. 1975. Therapy of advanced colorectal cancer with a combination of 5-fluorouracil, methyl-1,3-cis-(2-chloroethyl)-1-nitrosourea, and vincristine. *J. Natl. Cancer Inst.* 54:69–71
89. Carter, S. K., Friedmann, M. 1974. Integration of chemotherapy into combined modality treatment of solid tumors. II. Large bowel carcinoma. *Cancer Treatment Rev.* 1:111–29
90. Higgins, G. A. Jr., Humphrey, E., Juler, G. L., LeVeen, H. H., McCaughan, J., Keehn, R. J. 1976. Adjuvant chemotherapy in the surgical treatment of large bowel cancer. *Cancer* 38:1461–67

91. Dwight, R. W., Humphrey, E. W., Higgins, G. A., Keehn, R. J. 1973. FUDR as an adjuvant to surgery in cancer of the large bowel. *J. Surg. Oncol.* 5:243–49
92. Lawrence, W. Jr., Terz, J. J., Horsley, S. III, Donaldson, M., Lovett, W. L., Brown, P. W., Ruffner, B. W., Regelson, W. 1975. Chemotherapy as an adjuvant to surgery for colorectal cancer. *Ann. Surg.* 181:616–23
93. Grage, T. B. 1977. See Ref. 1, pp. 259–63
94. Li, M. C., Ross, S. T. 1977. See Ref. 1, pp. 273–77
95. Mavligit, G. M., Gutterman, J. U., Burgess, M. A., Khankhanian, N., Seibert, G. B., Speer, J. F., Jubert, A. V., Martin, R. C., McBride, C. M., Copeland, E. M., Gehan, E. A., Hersh, E. M. 1976. Prolongation of postoperative disease-free interval and survival in human colorectal cancer by BCG or BCG plus 5-fluorouracil. *Lancet* 1:871–75
96. Mavligit, G. M., Gutterman, J. U., Malahy, M. A., Burgess, M. A., McBride, C. M., Jubert, A., Hersh, E. M. 1977. See Ref. 1, pp. 279–86
97. Harvey, H., Badder, E., Demuth, W., Nahrwold, D., Patterson, L., Ricci, J., Shaver, W., White, D., Lipton, A. 1977. Oral 5-fluorouracil (5-FU) as an adjuvant to surgery in carcinoma of the colon. *Proc. Am. Assoc. Cancer Res.* 18:314 (Abstr.)
98. Comis, R. L., Carter, S. K. 1974. Integration of chemotherapy into combined modality treatment of solid tumors. III. Gastric cancer. *Cancer Treatment Rev.* 1:221–38
99. Carter, S. K., Comis, R. L. 1977. Gastric cancer. The current status of treatment. *J. Natl. Cancer Inst.* 58:567–78
100. Blokhina, N. G., Garin, A. M., Moroz, L. V., Pershin, M. P., Chumakina, S. I., Gunka, I. I., Pantophel, A. M., Steklenev, N. A., Zirin, M. A., Grjaznova, I. N., Shain, A. A., Pecherskaja, B. G., Saveliev, N. P., Bashirova, N. G., Kuzmin, V. P., Budarina, E. M., Svedencov, E. P., Phateeva, K. V., Iunusmetov, I. R., Silitrin, N. P., Shklovskii, G. S., Balas, A. N., Ushivceva, A. E. 1972. Treatment with 5-fluorouracil in prophylaxis of relapses and metastases of stomach cancer. *Neoplasma* 19:351–56
101. Fujimoto, S., Akao, T., Itoh, B., Kishiruka, I., Koyano, K., Kitsukawa, Y., Takahashi, M., Minami, T., Ishigami, H., Miyaraki, M., Amamiya, K., Ohyama, Y., Ono, K., Kure, M., Itoh, K., Hikosaka, T. 1977. Protracted oral chemotherapy with fluorinated pyrimidines as an adjuvant to surgical treatment for stomach cancer. *Ann. Surg.* 185:462–66
102. De Vita, V. T. Jr., Fisher, R. I. 1976. Natural history of malignant melanoma as related to therapy. *Cancer Treatment Rep.* 60:153–57
103. Gutterman, J. U., McBride, C., Freireich, E. J., Mavligit, G., Frei, E. III, Hersh, E. M. 1973. Active immunotherapy with BCG for recurrent malignant melanoma. *Lancet* 1:1208–12
104. Eilber, F. R., Morton, D. L., Holmes, E. L., Sparks, F. C., Ramming, K. P. 1976. Adjuvant immunotherapy with BCG in treatment of regional lymph-nodes metastases from malignant melanoma. *N. Engl. J. Med.* 294:237–40
105. Gutterman, J. U. 1977. Cancer systemic active immunotherapy today. Prospects for tomorrow. *Cancer Immunol. and Immunother.* 2:1–9
106. Hill, G. J., Metter, G. E., Uloss, S. E., Krementz, E., Fletcher, W., Minton, J., Golomb, F. 1976. DTIC therapy for melanoma: correlation of toxicity with response and longevity in 742 patients. *Proc. Am. Assoc. Cancer Res.* 17:244 (Abstr.)
107. Ikonopisov, R. C. 1975. The use of BCG in combined treatment of malignant melanoma. *Behring Inst. Mitt.* 56: 206–14
108. McPherson, T. A., Paterson, A. H. G., Williams, D., Watson, M. 1977. See Ref. 1, pp. 439–46
109. Jacquillat, C., Banset, P., Civatte, J., Puissant, A., Chastang, C. 1976. *Proc. Am. Assoc. Cancer Res.* 17:42 (Abstr.)
110. McMurtrey, M. J., Sinkovics, J. G., Plager, C., Romero, J. J., Romsdahl, M. M. 1977. Adjuvant chemoimmunotherapy for stage III (regional lymph nodes metastases malignant melanoma). *Proc. Am. Assoc. Cancer Res.* 18:308 (Abstr.)
111. Walker, M. D; 1975. Brain tumors chemotherapy: adjuvant to surgery

and radiation therapy. *Semin. Oncol.* 2:69–72

112. Band, P. R., Weir, B. K. H., Urtasun, R. C., Blain, G., McLean, D., Wilson, F., Mielke, B., Grace, M. 1974. Radiotherapy and CCNU in grade III and IV astrocytoma. *Proc. Am. Assoc. Cancer* Res. 14:161 (Abstr.)
113. Reagan, T. J., Bisel, H. F., Childs, D. S., Layton, D. D., Rhoton, A. L., Taylor, W. F. 1976. Controlled study of CCNU and radiation therapy in malignant astrocytoma. *J. Neurosurg.* 44:186–90
114. EORTC Brain Tumor Group. 1976. Effect of CCNU on survival, rate of objective remission and duration of free interval in patients with malignant brain glioma. First evaluation. *Eur. J. Cancer* 12:41–45
115. Walker, M. D., Strike, T. A. 1976. An evaluation of methyl CCNU, BCNU and radiotherapy in the treatment of malignant glioma. *Proc. Am. Assoc. Cancer Res.* 17:163 (Abstr.)
116. Gilbert, H. A., Sadoff, L., Eder, D., Cove, K., Wagner, J., Kagan, A. R., Nussbaum, H., Chan, P., Rao, A. 1977. See Ref. 1, pp. 487–94
117. DeCarvalho, S., Kaufman, A., Pineda, A. 1977. See Ref. 1, pp. 495–502
118. Carter, S. K., Wassermann, T. H. 1975. The chemotherapy of urologic cancer. *Cancer* 36:729–47
119. Samuels, M. L., Lanzotti, V. J., Holoye, P. Y., Eamonn Boyle, L., Smith, T. L., Johnson, D. E. 1976. Combination chemotherapy in germinal cell tumors. *Cancer Treatment Rev.* 3:185–204
120. Ansfield, F. J., Ramirez, G. 1977. See Ref. 1, pp. 321–26
121. Pommatau, E., Brunat, M., Chassard, J. L., Mayer, M. 1976. The value of chemotherapy in the treatment of testicular disembryoplastic tumors. *J. Surg. Oncol.* 8:211–15
122. Amiel, J. L., Cukier, J., Bacant, J., Beurton, D., Lemaire, Ph., Rouesse, J., Droz, J. P. 1976. Les traitment actual des cancers du testicule. *Nouv. Presse Med.* 5:2929–32
123. Johnson, D. E., Bracken, R. B., Ayola, A. G., Samuels, M. L. 1976. Retroperitoneal lymphadenectomy as adjunctive therapy in selected cases of advanced testicular carcinoma. *J. Urol.* 116:66–68
124. Merrin, C., Takita, H., Weber, R., Waisman, Z., Baumgartner, G., Murphy, G. P. 1976. Combination radical surgery and multiple sequential chemotherapy for the treatment of advanced carcinoma of the testis (stage III). *Cancer* 37:20–29
125. Fisher, R. I., Young, R. C. 1977. Advances in the staging and treatment of ovarian cancer. *Cancer* 39:967–72
126. Griffiths, C. T., Grogan, R. H., Hall, T. C. 1972. Advanced ovarian cancer: primary treatment with surgery, radiotherapy, and chemotherapy. *Cancer* 29:1–7
127. Julian, C. G., Woodruff, J. D. 1969. The role of chemotherapy in the treatment of primary ovarian malignancy. *Obstet. Gynecol. Surv.* 24:1307–42
128. Smith, J. P., Rutledge, F. 1975. Advances in chemotherapy for gynecologic cancer. *Cancer* 36:669–74
129. Hreshchyshyn, M. M. 1975. The results of the Gynecologic Oncology Group trials on ovarian cancer: preliminary report. *Natl. Cancer Inst. Monogr.* 42:155–65
130. Hreshchyshyn, M. M., Norris, H. J. 1977. Postoperative treatment of women with resectable ovarian cancer with radiotherapy, melphalan, or no further treatment. *Proc. Am. Assoc. Cancer Res.* 18:195 (Abstr.)
131. Morton, D. L., Eilber, F. R., Townsend, C. M., Grant, T. T., Mirra, J., Weisenburger, T. H. 1976. Limb salvage from a multidisciplinary treatment approach for skeletal and soft tissue sarcomas of the extremity. *Ann. Surg.* 184:268–77
132. Morton, D. L., Eilber, F. R., Holmes, E. C., Townsend, C. M., Mirra, J., Weisenburger, T. H. 1977. See Ref. 1, pp. 391–97
133. Maurer, H. M., for IRS Committee. 1976. Intergroup rhabdomyosarcoma study: Progress report. *Proc. Am. Assoc. Cancer Res.* 17:242 (Abstr.)
134. Goldsmith, M. A., Carter, S. K. 1975. The integration of chemotherapy into a combined modality approach to cancer therapy. V. Squamous cell cancer of the head

and neck. *Cancer Treatment Rev.* 2:137–58

135. Taylor, S. G. IV, Bytell, D. E., Sisson, G. A., Nisins, S., De Wys, W. D. 1977. See Ref. 1, pp. 457–66

136. Durie, B. G. M., Hellman, D., O'Mara, R. E., Woolfenden, J., Kartchner, M., Salmon, S. E. 1977. See Ref. 1, pp. 475–86

137. Lewis, G. C., Slack, N. H., Mortel, R., Bross, I. K. J. 1974. Adjuvant progesten therapy in primitive definitive treatment of endometrial carcinoma. *Gynecol. Oncol.* 2:368–76

Ann. Rev. Med. 1978. 29:177–89
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CLINICAL HYPOPHOSPHATEMIA ◆7275

Faith Fitzgerald, M.D.
Medical Service, San Francisco General Hospital, and Department of Medicine,
University of California, San Francisco 94143

INTRODUCTION

Some older clinicians view the recent flurry of attention given to phosphate deficiency with a bemused satisfaction. They knew 40 years ago that serum phosphate could drop precipitously in diabetic ketoacidosis, and they supplemented this ion as a regular part of the therapy of that disease. It seemed the reasonable thing to do: Phosphorus is a major intracellular anion in mammals, essential to the structural integrity of cells and to a wealth of basic metabolic processes.

Then, for some 25 years, we clinicians seemed to forget phosphate deficiency. Certainly, we used phosphorus measurements in the assessment of calcium homeostasis. We vigorously treated the rise of serum phosphorus in renal disease. But only in the past decade has knowledge of phosphorus deficiency again emerged from the province of an arcane academic exercise into the mainstream of practical, clinically pertinent information.

Modern vigorous therapies, such as hemodialysis and hyperalimentation, may induce profound phosphate deficiency. Automated multichannel blood chemical analyses are now widely used, and laboratories report phosphorus levels to us even though we may not have thought to ask specifically for them. These two events, the more frequent induction and the laboratory recognition of phosphate deficiency, have led to our awareness of an increasing number of morbid clinical consequences of this deficit.

PHOSPHORUS METABOLISM

The adult human body contains between 500 and 700 g of phosphate, about 85% of which is in bone. The rest is located primarily intracellularly, where it functions organically as a structural component of ribonucleic acids and of the phospholipid membranes of cells and organelles. Phosphorylated adenine and guanine nucleotides are essential to the energy systems of intracellular work. Inorganic phosphorus participates in the regulation of the rate of glycolysis, hence the production of energy.

0066-4219/78/0401-0177$01.00

Dietary intake varies enormously, but it is seldom deficient since phosphorus is abundant in ordinary foods (1, 2) and is readily absorbed by the normal intestine. The kidney is the principal excretory route, and the complexities of the renal handling of phosphorus by glomerular filtration, tubular secretion, and hormonal controls are still under intensive study.

Serum phosphorus levels in adults range between 3.0 and 4.5 mg%. These normal concentrations may vary slightly with diet, time of day, respiratory rate, and hormonal status. Young children have somewhat higher normal values (5–7 mg%).

Symptoms attributable to hypophosphatemia have been described when levels of serum phosphorus fall to 2 mg% or less, with the most disturbing morbidity at levels of 0.2% and below. The individual physician's chances of encountering clinically significant hypophosphatemia clearly depend upon the nature of his practice. In an Australian general medical hospital, a survey showed 2% of all admitted patients to have a serum phosphorus of less than 2 mg% (3); in an American Veterans Administration Hospital sample, some 10% of admissions for complications of alcoholism had a serum level of less than 2 mg% (4).

CAUSES OF HYPOPHOSPHATEMIA

Table 1 summarizes the principal causes of hypophosphatemia so far described. These causes are discussed in more detail below.

Decreased Intake

The ubiquity of sufficient phosphorus in foodstuffs makes selective dietary deficiency of this essential nutrient extremely rare in nature (2). But the patient who is starving, malabsorbing, or vomiting may be compromised, and this no doubt plays a major role in the frequency of hypophosphatemia in the down-and-out alcoholic population seen in many county hospitals. Chronic ingestion of aluminum or magnesium hydroxide antacids may lead to hypophosphatemia, since these compounds bind phosphate in the gut, which results in selective malabsorption (5–11). In a way, hyperalimentation with phosphate-poor solutions may be seen as analogous to selective phosphate starvation, though transcellular phosphate shifts probably play the major role in the hypophosphatemia of the hyperalimented patient (12–25). As hospitals become more familiar with the requirements of total parenteral nutrition, the incidence of unanticipated hyperalimentation hypophosphatemia should decline.

Transcellular Shifts

The administration of carbohydrate lowers serum phosphate. There is a normal postcibal serum phosphate fall of about 0.25 mg% (26). This drop is caused by the movement of phosphorus with glucose into cells (26, 27). It may be markedly accentuated in patients given parenteral glucose (28), especially if they are starving (29). It may be that less of the administered carbohydrate in malnourished patients goes to liver storage and more to immediate muscle needs where most phosphorus is utilized (29). This seems to be the major mechanism of profound hypophosphatemia in the patient given carbohydrate-rich hyperalimentation.

The practical clinical implication derived from this is that the phosphorus measured in the serum of malnourished patients upon entry to the hospital may be markedly higher than that measured again after the administration of parenteral dextrose solutions. Since the passage of glucose, and thus of phosphorus, into cells is facilitated by insulin (27) and alkalosis (30–32), the serum phosphorus level may likewise plummet when the diabetic patient in ketoacidosis is first treated. I strongly recommend, therefore, that these classes of patients have serial phosphate determinations done rather than the more usual single admission value.

Of the other causes of "shift hypophosphatemia" listed in Table 1, those associated with alkalosis (30–32) and liver disease (32–41) may be of clinical significance, and are discussed shortly.

Increased Loss

Any form of renal tubular phosphate leak may ultimately lead to hypophosphatemia. Diuretics, including osmotic diuretics, may increase phosphate clearance (42, 43). Concurrent hypokalemia (39–41) and hypomagnesemia (4), as may occur in alcoholism or diabetes (44, 45), impair renal retention of phosphorus. Acidemic states, by presenting more phosphorus to the glo-

Table 1 Causes of hypophosphatemia

Decreased intake		
Diet	Starvation[a]	Antacids[a]
Malabsorption	Vomiting[a]	Hyperalimentation[a]
Transcellular shifts		
Carbohydrate (especially parenteral)[a]		Sepsis
Alkalosis[a]		Acute myocardial infarction
Liver disease (?[a])		Androgens/estrogens
Pregnancy		Catecholamines
Hypothyroidism		
Increased loss		
Hemodialysis (phosphate-poor bath)[a]		Hypomagnesemia (?[a])
Hypokalemia (?[a])		Gout
Diuretics		Renal tubular acidosis
Acidemia		Tumor phosphaturia
Genetic hypophosphatemia		
Multiple mechanisms		
Alcoholism[a]		Diabetic ketoacidosis[a]
Hyperparathyroidism/Vitamin D abnormalities[a]		

[a] Most clinically important causes in terms of frequency or severity.

merulus, increase its excretion (46, 47). Rarely, a tumor may, through mechanisms as yet not known but thought to be hormonal, cause renal tubular phosphorus wasting (48, 49). Perhaps the leading cause of phosphate deficiency syndromes in this category is that of hemodialysis against a phosphate-poor bath (9–11). Inadequate supplementation of phosphate in the dialysis solutions may be aggravated by the not infrequent concurrence of malnutrition in the uremic patient and the regular use of phosphate-binding antacids.

MULTIPLE MECHANISMS

The alcoholic patient may be starving, vomiting, or malabsorbing from his pancreatic disease (3). His frequent concurrent hypokalemia and hypomagnesemia may increase renal phosphate loss (4, 39–41). The catecholamine excess (3, 28) of withdrawal, the alkalosis of hyperventilation (30–31), his underlying liver disease (32–34), his sepsis (3, 50, 51), and his frequent receipt of antacid and carbohydrate therapy in the hospital all conspire to make him a victim of clinically significant hypophosphatemia. The diabetic ketoacidotic (44, 45) who is vomiting and anorexic may lose significant amounts of phosphorus from his glycosuric diuresis, hypokalemia, and acidemia. Then, in the hospital, the patient is given insulin, bicarbonate, and carbohydrate. Hypophosphatemia is an anticipated result. The patient with hyperparathyroidism loses phosphate through the kidneys, may not eat, or may be given antacids for his hypercalcemic ulcer disease. And, while hyperparathyroidism alone is seldom associated with levels of serum phosphorus less than 1.5 mg% (53), some of the symptoms of his disease may be attributable to hypophosphatemia (5) rather than to the more commonly considered calcium elevations.

CLINICAL SYNDROMES OF HYPOPHOSPHATEMIA

To date, four major kinds of syndromic dysfunction have been discovered in the hypophosphatemic patient. These are outlined in Table 2.

NEUROMUSCULAR DYSFUNCTION

Normal muscular function depends upon ATP and organophosphorus compounds providing energy to the actin-myosin interaction and structural integrity to the muscle cells themselves (52). The rate of energy generation and utilization in muscle is, through glycogenolysis and glycolysis, likewise phosphorus dependent. The adequate oxygenation of working tissue requires that the red cells delivering oxygen be both ATP and 2,3-diphosphoglycerate sufficient (13). It is not surprising, therefore, that phosphorus deficiency in man has been associated with neuromuscular dysfunction. What is somewhat surprising is the association of this dysfunction with relatively minimal phosphate depressions in serum. Lotz (5) reported weakness, malaise, anorexia, bone pain, joint stiffness, and intention tremors in patients with antacid hypophosphatemia at levels of 2 mg%

or less. It is possible, therefore, that some of the discomfort of hyperparathyroidism is hypophosphatemic in origin. More profound hypophosphatemia, as with hemodialysis, has apparently resulted in such marked disturbances of neuromuscular function as anisocoria, paresthesias, seizures, coma, paralysis, and death (11, 17, 24). It has been suggested that rhabdomyolysis, especially in alcoholic patients, might in part be due to the loss of structural integrity of muscle cells induced by severe phosphate depletion (53).

HEMATOLOGIC SYNDROMES

The ease of hematologic bioassay has made this tissue the most fruitful in investigations of hypophosphatemia. Red cells, white cells, and platelets all require phosphorus for structure and function. Clinically significant deficiency syndromes have been proven or strongly suggested for all three.

Erythrocytes

The phosphorus incorporated into red cell ATP is derived directly from the extracellular inorganic serum phosphorus bathing that red cell. The rate of

Table 2 Clinical consequences of hypophosphatemia

Hematologic

- Erythrocytes:
  - Hemolytic anemia
  - Decreased oxygen release
- Polymorphonuclear leukocytes:
  - Decreased chemotaxis
  - Decreased phagocytosis
  - Decreased killing
- Platelets:
  - Decreased clot retraction
  - Decreased platelet survival

Hepatic

- ?Hepatic hypoxia (decreased red cell 2,3-DPG)
- ?Correlation with hepatic coma

Neuromuscular

- Muscle weakness
- Rhabdomyolysis
- Intention tremor
- Paresthesias
- Paralysis
- Coma
- Convulsions
- Anisocoria
- Ballismus
- Ataxia
- Hyporeflexia
- Abnormal EEG
- Abnormal EMG
- Abnormal nerve conduction

Skeletal

- Osteomalacia
- Arthritis

anaerobic metabolism of glucose in red cells can be shown to vary directly with the serum phosphorus level (54, 55). Thus the production of ATP by red cell glycolysis depends directly on the serum phosphorus level (56).

Red cell membranes contain structural phospholipids; moreover, adequate amounts of ATP are necessary to maintain membrane integrity and deformability (57, 58). When red cell ATP levels fall below 15% of their normal concentration, the erythrocyte membrane becomes more rigid, less plastic, and assumes a microspherocytic morphology (58, 59). Severe hemolytic anemias have been documented in patients hypophosphatemic to levels of 0.2 mg% and below (59, 60). In these patients, red cell ATP was measured at 11% of normal concentration and red cell survival by Cr^{51} tagging was only one fifth of normal (59). Presumably, lesser degrees of hypophosphatemia might contribute to hemolysis in patients with other insults to red cell structure, as in those with alcoholism, sepsis, or uremia.

But even if structural integrity were maintained, the red cell would be functionally compromised by hypophosphatemia. The concentration of red cell 2,3-diphosphoglycerate is a major determinant of the avidity of binding of oxygen to hemoglobin: The higher the erythrocytic content of 2,3-DPG, the more easily does hemoglobin surrender its oxygen to tissues (12, 61). In patients rendered hypophosphatemic by phosphate-poor hyperalimentation, there is a measurable decrease in red blood cell 2,3-DPG (12, 13) with a consequent shift to the left of the oxyhemoglobin dissociation curve (12, 13, 22, 59, 62) and decreased delivery of oxygen to tissues.

In hypophosphatemic patients, then, it is possible that some of the neuromuscular symptoms are simply those of tissue hypoxia. This has clinically obvious pertinence, especially in acidemic patients—uremics or diabetic ketoacidotics—to whom the clinician may be tempted to give large amounts of bicarbonate. Acidemia shifts the oxyhemoglobin dissociation to the right, with subsequent increase of tissue oxygenation (63, 64). Alkali not only decreases serum phosphate, and thus erythrocytic 2,3-DPG, but independently shifts oxyhemoglobin dissociation to the left.

Leukocytes

Hyperalimented patients are subject to episodes of severe sepsis with bacteria and fungi. This has been attributed to the prolonged presence of an indwelling line infusing rich medium into an immunosuppressed host. Craddock and his colleagues (21) investigated the role of hypophosphatemia in this fearful complication of total parenteral nutrition, and documented a marked decrease in incidence of infection if the patients were phosphate sufficient.

Hyperalimented hypophosphatemic dogs showed severe compromise in leukocyte chemotaxis, phagocytosis, and killing, with a concurrent decrease in white cell ATP. However, polymorphonuclear leukocytes supplemented in vivo with phosphate, or in vitro with adenosine and phosphate, restored their function (21).

ATP is required for the formation of phagocytic vacuoles in white cells (21), and provides the energy for membrane work, amoeboid movement, and pseudopod formation (65, 66). Organophosphorus compounds are building blocks for membrane synthesis and phagocytic granules (67). When serum phosphorus drops below 1 mg%, the deficiency becomes functionally significant and white cell compromise ensues (21). One can speculate that the same phenomenon may contribute to the known predisposition to infection of diabetics, alcoholics, and others who have a tendency toward acquired hypophosphatemia.

Clinical observations have shown some correlation of hypophosphatemia with septic states—both gram-negative and gram-positive bacteremias (3, 50, 51). The question remains whether the sepsis causes the drop in serum phosphate or vice versa.

Platelets

Hypophosphatemic patients have been said to show an increased tendency to epistaxis and mild gastrointestinal bleeding (52). Yawata et al (19), working with hypophosphatemic dogs, demonstrated decreased platelet ATP, with absent clot retraction and decreased platelet survival times. Hypophosphatemia may therefore potentially compound the thrombopathy of alcoholism or uremia; patients may become hypophosphatemic through the prolonged use of phosphate-binding antacids given *because* they are bleeding. Bank blood, if stored in phosphate-poor solutions, may be platelet and 2,3-DPG deficient (68). It is important, therefore, to pay more attention to the phosphate levels of our bleeding patients.

HEPATIC DISEASE

The role of phosphate in hepatic decompensation remains conjectural (33–37). Rajan (34) observed that hypophosphatemic alcoholics tended to undergo more rapid worsening of their liver function in hospital than did the normophosphatemic. He proposed that hypophosphatemia might blunt the normal compensatory rise in red cell 2,3-DPG level that occurs in liver disease. If this were so, the liver—already compromised by shunting of its portal flow—would receive even less oxygenation. Do we, by vigorously treating our alcoholics with parenteral glucose, antacids, and diuretics, precipitate a hypophosphatemic worsening of their liver disease?

Others have argued that the hypophosphatemia of liver disease may be the consequence of hepatic dysfunction rather than its cause. The respiratory alkalosis that accompanies liver failure will shift phosphate into cells (31); or the inability of the cirrhotic liver to lay down glycogen may shift prosphate to muscle, where more is utilized, leading to even greater negative balance in the already phosphate-deficient patient (29).

It appears that the alcoholic patient, whose starvation, vomiting, malabsorption, and electrolyte deficiencies predispose him to hypophosphatemia, may suffer far more commonly than we realize from its multiple consequences: Hemolytic anemias, white cell dysfunction, bleeding diatheses, rhabdomyolysis, confusional states, and hepatic disease are, even if not solely attributable to hypophosphatemia, potentially subject to accentuation by phosphate deficiency in the alcoholic.

SKELETAL EFFECTS

Early studies in nutrition showed that artificially induced phosphate deficiency in animals led to clinical bone disease resembling rickets (69–71). This effect of phosphate deficiency was independent of either vitamin D or parathyroid hormone (72, 73).

A clinical parallel to the animal studies was found in man when it was noted that several patients in long-term therapy with phosphate-binding anticids developed osteomalacia of severity sufficient to give pathologic fractures (6, 9). More recently, Moser & Fessel (49) have described a phosphate deficiency syndrome of large-joint arthralgias and inflammatory arthritis, with prominent sacroiliitis so severe as to be initially misdiagnosed as ankylosing spondylitis.

Studies on these effects of phosphate deficiency, as in the patients with the rare genetic hypophosphatemia, have been variable and inconclusive with regard to the role of parathyroid hormone in mediating the skeletal disease (74).

PHYSIOLOGIC RESPONSE TO INDUCED HYPOPHOSPHATEMIA

Lotz and his colleagues (5) studied a series of patients in whom prolonged use of phosphate-binding antacids had led to hypophosphatemia. They described the physiologic responses of these patients to phosphate depletion:

1. *Rapid decrease in urinary phosphorus* In those patients in whom renal tubular dysfunction is not causative or contributory, one may anticipate an early and almost total renal conservation of phosphorus even with very small losses relative to the total body phosphorus stores.

Since the renal conservation of phosphorus exceeds that expected from decreased glomerular delivery alone, a "phosphorus-restraining hormone" has been postulated (5). It may be that intratubular cell phosphate is directly regulatory. How this mechanism fits in with the renal metabolic interactions of vitamin D-, parathyroid hormone-, and renal tubular cell phosphorus remains conjectural (38, 39). However, the physiologic responses to hypophosphatemia occur, seemingly unaltered, in the parathyroidectomized patient.

2. *Hypercalciuria* Hypophosphatemic patients increase renal excretion of calcium up to six times the normal level, a calcium wastage sufficient to produce a negative total body calcium balance. The reabsorption of bone, which no doubt occurs to provide for this marked calcium excretion, may be a major mechanism of the osteomalacia of hypophosphatemia. In spite of the hyper-

calciuria, renal stone disease has not yet been shown to be increased in the hypophosphatemic patient.

The mechanism of the increased calcium excretion is unknown (5, 75). It occurs, as do the other physiologic responses to phosphate depletion, in the surgically hypoparathyroid patient.

3. *Increased GI absorption of calcium* Increased GI absorption of calcium is insufficient to compensate for renal calcium wastage, and the severely hypophosphatemic patient is often depleted in total body calcium. This calcium depletion assumes clinical importance when phosphorus is too rapidly repleted: A reciprocal fall in calcium may be induced as serum phosphorus is normalized, which leads to tetany (7, 11).

THERAPY

The vigor of therapy should, as with any disorder, depend on the severity of the disease. It is often difficult, however, for the clinician to distinguish between the manifestations of hypophosphatemia (anemia, infection, confusion, bleeding, aches, and pains) and those of the concurrent medical condition that induced it (starvation, alcoholism, diabetes, etc).

In the vast majority of patients, who are either without symptoms or only mildly symptomatic, the best therapy is simple discontinuation or avoidance of those factors that induce the hypophosphatemia. The patient will commonly spontaneously elevate his serum phosphorus to normal levels over the next several days (4, 9, 49, 76).

In patients with symptomatic, but not life-threatening, hypophosphatemia (as in arthritis, weakness, paresthesias, mild confusion, or infection), one may feel compelled to supplement dietary phosphorus. If the patient is able to eat, dairy products are both well accepted to patient taste and advantageous in that they supply absorbable calcium and phosphorus (Thus they avoid the hypocalcemia which phosphorus alone might induce.). In more severe circumstances of hypophosphatemia, supplemental oral phosphorus, as the Fleets® phosphosoda (15–30 ml orally 3 times a day) may be given (76). One must monitor the calcium with this regimen, and be prepared for patient complaints of diarrhea as an unpleasant side effect of therapy.

In profound hypophosphatemia, with hemolytic anemia, coma, and seizure, or paralysis, the clinician may be compelled to administer parenteral phosphate. Few guidelines exist, but monobasic potassium phosphate, 60 meq intravenously 3 times a day, has been used with success (10). In this regimen, the patient is at risk for hyperkalemia, hyperphosphatemia, and hypocalcemia, and must be assiduously monitored.

Hypophosphatemia is not an endocrinologic biochemist's esoteric hobby. Rather, as may be seen from the flurry of case reports in the literature over recent years, it is an increasingly recognized cause of morbidity and potential mortality in patients with a number of other more or less serious disorders. Thus, although the bench investigation of hypophosphatemia is of immense satisfaction to the scientist confronted with the thorny problem of why so ubiquitous and vital an element is so subject to depletion in so many circumstances, it is also of immediate concern for clinicians and their patients to recognize and remedy hypophosphatemia, particularly since it is so frequently iatrogenically induced.

Literature Cited

1. Follis, R. H. Jr. 1966. Rickets and osteomalacia. In *Principles of Internal Medicine,* ed. T. R. Harrison, pp. 383–89. New York: McGraw-Hill. 1874 pp. 5th ed.
2. National Research Council, Food and Nutrition Board. 1964. *Recommended Dietary Allowances: A Report.* 6th ed. Washington DC
3. Betro, M. G., Pain, R. W. 1972. Hypophosphatemia in a hospital population. *Br. Med. J.* 1:273–76
4. Stein, J. H., Smith, W. O., Ginn, H. E. 1966. Hypophosphatemia in acute alcolism. *Am. J. Med. Sci.* 252:78–83
5. Lotz, M., Zisman, E., Bartter, F. C. 1968. Evidence for a phosphorus-depletion syndrome in man. *N. Engl. J. Med.* 278:409–15
6. Bloom, W. L., Flinchum, D. 1960. Osteomalacia with pseudofractures caused by the ingestion of aluminum hydroxide. *J. Am. Med. Assoc.* 174:1327–30
7. Lotz, M., Ney, R., Bartter, F. C. 1964. Osteomalacia and debility resulting from phosphorus depletion. *Trans. Assoc. Am. Physicians* 77:281–95
8. Ludwig, G. D., Kyle, G. C., de Blanco, M. 1967. "Tertiary" hyperparathyroidism induced by osteomalacia resulting from phosphorus depletion. *Am. J. Med.* 43:136–40
9. Abrams, D. E., Silcott, R. B., Terry, R., Berne, T. V. 1974. Antacid induction of phosphate depletion syndrome in renal failure. *West. J. Med.* 120:157–60
10. Lichtman, M. A., Miller, D. R., Freeman, R. B. 1969. Erythrocyte adenosine triphosphate depletion during hypophosphatemia in a uremic subject. *N. Engl. J. Med.* 280:240–44
11. Boelens, P. A., Norwood, W. K., Kjellstrand, C., Brown, D. M. 1970. Hypophosphatemia with muscle weakness due to antacids and hemodialysis. *Am. J. Dis. Child.* 120:350–53
12. Lichtman, M. A., Miller, D. R., Cohen, J., Waterhouse, C. 1971. Reduced red cell glycolysis, 2,3-diphosphoglycerate and adenosine triphosphate concentration, and increased hemoglobin-oxygen affinity caused by hypophosphatemia. *Ann. Intern. Med.* 74:562–68
13. Travis, S. F., Sugerman, H. J., Ruberg, R. L., Dudrick, S. J., Delivoria-Papadopoulos, M., Miller, L. D., Oski, F. A. 1971. Alterations of red-cell glycolytic intermediates and oxygen transport as a consequence of hypophosphatemia in patients receiving intravenous hyperalimentation. *N. Engl. J. Med.* 285:763–68
14. Metzger, R., Burke, P., Thompson, A., Lordon, R., Frimpter, G. W. 1971. Hypophosphatemia and hypouricemia during parenteral hyperalimentation with an amino acid–glucose preparation. *J. Clin. Invest.* 50:65a–66a
15. Allen, T. R., Ruberg, R. L., Dudrick, S. J., Long, J. M., Steiger, E. 1971. Hypophosphatemia occurring in patients receiving total parenteral hyperalimentation. *Fed. Proc.* 30:580
16. Sugerman, H., Travis, S., Pollock, T., Ruberg, R., Dudrick, S., Delivoria-Papadopoulos, M., Miller, L., Oski, F. 1971. Alterations in oxygen transport and red cell metabolism as a consequence of hypophosphatemia in intravenous hyperalimentation. *Clin. Res.* 19:487
17. Silvis, S. E., Paragas, P. D. Jr. 1972. Paresthesia, weakness, seizures, and hypophosphatemia in patients receiving hyperalimentation. *Gastroenterology* 62:513–20
18. Dudrick, S. J., MacFadyen, B. V. Jr., Van Buren, C. T., Ruberg, R. L., Maynard, A. T. 1972. Parenteral hyperalimentation: metabolic problems and solutions. *Ann. Surg.* 176:259–64
19. Yawata, Y., Craddock, P., Hebbel,

R., Howe, R., Silvis, S., Jacob, H. 1973. Hyperalimentation hypophosphatemia: hematologic-neurologic dysfunction due to ATP depletion. *Clin. Res.* 21:729
20. Craddock, P. R., Yawata, Y., Silvis, S., Jacob, H. 1973. Phagocyte dysfunction induced by intravenous hyperalimentation. *Clin. Res.* 21:597
21. Craddock, P. R., Yawata, Y., VanSanten, L., Gilbertstadt, S., Silvis, S., Jacob, H. S. 1974. Acquired phagocytic dysfunction: a complication of the hypophosphatemia of parenteral hyperalimentation. *N. Engl. J. Med.* 290:1403–7
22. Lichtman, M. A. 1974. Hypoalimentation during hyperalimentation. *N. Engl. J. Med.* 290:1432–33
23. Prins, J. C., Schrijver, H., Staghouwer, J. M. 1973. Hyperalimentation, hypophosphatemia, and coma. *Lancet* 1:1253–54
24. Weintraub, M. I. 1976. Hypophosphatemia mimicking acute Guillain-Barre-Strohl syndrome: a complication of parenteral hyperalimentation. *Am. J. Med. Assoc.* 235:1040–41
25. Carruthers, B. M., Copp, D. H., McIntosh, H. W. 1964. Diurnal variation in urinary excretion of calcium and phosphate and its relation to blood levels. *J. Lab. Clin. Med.* 63:959–68
26. Annino, J. S., Relman, A. S. 1959. The effect of eating on some of the clinically important chemical constituents of the blood. *Am. J. Clin. Pathol.* 31:155–59
27. Forsham, P. H., Thorn, G. W. 1950. Changes in inorganic serum phosphorus during intravenous glucose tolerance test as adjunct to diagnosis of early diabetes mellitus. *Proc. Am. Diabetes Assoc.* 9:101–22
28. Danowski, T. S., Gillespie, H. K., Fergus, E. B., Puntereri, A. J. 1956. Significance of blood sugar and serum electrolyte changes in cirrhosis following glucose, insulin, glucagon, or epinephrine. *Yale J. Biol. Med.* 29:361–75
29. Corredor, D. G., Sabeh, G., Mendelsohn, L. V., Wasserman, R. E., Sunderman, J. H., Danowski, T. S. 1969. Enhanced postglucose hypophosphatemia during starvation therapy of obesity. *Metabolism* 18:754–63
30. Okel, B. B., Hurst, J. W. 1961. Prolonged hyperventilation in man: associated electrolyte changes and subjective symptoms. *Arch. Intern. Med.* 108:757–62
31. Mostellar, M. E., Tuttle, E. P. Jr. 1964. Effects of alkalosis on plasma concentration and urinary excretion of inorganic phosphate in man. *J. Clin. Invest.* 43:138–49
32. Jacob, H. S. 1975. Severe hypophosphatemia: a previously ignored cause of cellular dysfunction. *West. J. Med.* 122:501–2
33. Amatuzio, D. S., Stutzman, F., Shrifter, N., Nesbitt, S. 1952. Study of serum electrolytes (Na, K, Ca, P) in patients with severely decompensated portal cirrhosis of liver. *J. Lab. Clin. Med.* 39:26–29
34. Rajan, K. S., Levinson, R., Leevy, C. M. 1973. Hepatic anoxia secondary to hypophosphatemia. *Clin. Res.* 21:521
35. Roberts, K. E., Vanamee, P., Poppell, J. W., Rubin, A., Braveman, W., Randall, H. T. 1956. Electrolyte alterations in liver disease and hepatic coma. *Med. Clin. North Am.* 40:901–14
36. Frank, B. W., Kern, F. Jr. 1962. Serum inorganic phosphorus during hepatic coma. *Arch. Intern. Med.* 110:865–71
37. Keating, J. P., Karl, I. E., DeVivo, D. C., Haymond, M. W. 1975. Hypophosphataemia in Reye's syndrome. *Lancet* 2:39–40
38. Tanaka, Y., DeLuca, H. F. 1973. The control of 25-hydroxy vitamin D metabolism by inorganic phosphorus. *Arch. Biochem. Biophys.* 154:566–74
39. Sawyer, W. T. 1974. Phosphorus depletion and vitamin D metabolism (cont.). *N. Engl. J. Med.* 290:345
40. Anderson, D. C., Peters, T. J., Stewart, W. K. 1969. Association of hypokalaemia and hypophosphatemia. *Br. Med. J.* 4:402–3
41. Vianna, M. J. 1971. Severe hypophosphatemia due to hypokalemia. *J. Am. Med. Assoc.* 215:1497–98
42. Duarte, C. G. 1968. Effects of chlorothiazide and amipramazide (MK870) on the renal excretion of calcium, phosphate, and magnesium. *Metabolism* 17:420–29
43. Duarte, C. G., Bland, J. H. 1965. Calcium, phosphorus and uric acid clearances after intravenous admin-

istration of chlorothiazide. *Metabolism* 14:211–19

44. Alberti, K. G. M. M., Darley, J. H., Emerson, P., Hockaday, T. D. R. 1972. 2,3-Diphosphoglycerate and tissue oxygenation in uncontrolled diabetes mellitus. *Lancet* 2:391–95
45. Guest, G. M. 1942. Organic phosphate of blood and mineral metabolism in diabetic acidosis; president's address. *Am. J. Dis. Child.* 64:401–12
46. Haldane, J. B. S., Wigglesworth, D. B., Woodrow, C. E. 1924. The effect of reaction changes on human inorganic metabolism. *Proc. R. Soc. London Ser. B* 96:1
47. Guest, G. M., Rapoport, S. 1941. Organic acid-soluble phosphorus compounds of the blood. *Physiol. Rev.* 21:410–37
48. Olefsky, J., Kempson, R., Jones, H., Reaven, G. 1972. "Tertiary" hyperparathyroidism and apparent "cure" of vitamin-D-resistant rickets after removal of an ossifying mesenchymal tumor of the pharynx. *N. Engl. J. Med.* 286:740–45
49. Moser, C. R., Fessel, W. J. 1974. Rheumatic manifestations of hypophosphatemia. *Arch. Intern. Med.* 134:674–78
50. Riedler, G. F. 1972. Platelet count, white blood cell differential count, and inorganic phosphate: 3 useful laboratory signs in septicemia. *Schweiz. Med. Wochenschr.* 102:497–504
51. Riedler, G. F., Scheitlin, W. A. 1969. Hypophosphatemia in septicemia: higher incidence in gram-negative than in gram-positive infections. *Br. Med. J.* 1:753–56
52. Mommaerts, W. F. 1969. Energetics of muscular contraction. *Physiol. Rev.* 49:427–508
53. Knochel, J. 1977. The pathophysiology and clinical characteristics of severe hypophosphatemia. *Arch. Intern. Med.* 137:203–20
54. Schner, S. L. 1967. ATP synthesis in human erythrocyte membranes. *Biochim. Biophys. Acta* 135:591
55. Guest, G. M., Rapoport, S. 1939. Role of acid-soluble phosphorus compounds in red blood cells in experimental rickets, renal insufficiency, pyloric obstruction, gastroenteritis, ammonium chloride acidosis and diabetic acidosis. *Am. J. Dis. Child.* 58: 1072–89
56. Rose, I. A., Warms, J. V. 1966. Control of glycolysis in the human red blood cell. *J. Biol. Chem.* 241:4848–54
57. Nakao, M. Nakao, T., Yamazoe, S., Yoshikawa, H. 1961. Adenosine triphosphate and shape of erythrocytes. *J. Biochem. Tokyo* 49:487–92
58. Weed, R. I., LaCelle, P. L., Merrill, E. W. 1969. Metabolic dependence on red cell deformability. *J. Clin. Invest.* 48:795–809
59. Jacob, H. S., Amsden, T. 1971. Acute hemolytic anemia with rigid red cells in hypophosphatemia. *N. Engl. J. Med.* 285:1446–50
60. Klock, J. C., Williams, H. E., Mentzer, W. C. 1974. Hemolytic anemia and somatic cell dysfunction in severe hypophosphatemia. *Arch. Intern. Med.* 134:360–64
61. Bunn, H. F., Jandl, J. H. 1970. Control of hemoglobin function within the red cell. *N. Engl. J. Med.* 282:1414–21
62. Chanutin, A., Curnish, R. R. 1967. Effect of organic and inorganic phosphates on the oxygen equilibrium of human erythrocytes. *Arch. Biochem.* 121:96–102
63. Thomas, H. M., Lefrak, S. S., Irwin, R. S., Fritts, H. W. Jr., Caldwell, P. R. B. 1974. The oxyhemoglobin dissociation curve in health and disease. Role of 2,3-diphosphoglycerate. *Am. J. Med.* 57:331–48
64. Klock, R. A. 1972. Oxygen transport and 2,3-diphosphoglycerate (DPG). *Chest* 62 (Suppl.): 79S-85S
65. Pollard, T. D., Weihing, R. R. 1974. Actin and myosin and cell movement. *Crit. Rev. Biochem.* 2:1
66. Tatsumi, N., Shibata, N., Okamura, Y., Takeuchi, K., Senda, N. 1973. Actin and myosin A. from leukocytes. *Biochim. Biophys. Acta* 305:433–44
67. Tou, J. S., Stjernholm, R. L. 1974. Stimulation of the incorporation of ^{32}Pi and *myo*[2-^{3}H]inositol into the phosphoinositides in polymorphonuclear leukocytes during phagocytosis. *Arch. Biochem. Biophys.* 160:487–94
68. Huisman, T. H., Boyd, E. M., Kitchens, J., Mayson, S., Shepeard, W. L. 1969. Oxygen equilibrium and biochemical changes of whole

blood stored in different preservation media. *Transfusion* 9:180–90

69. Aubel, C. E., Hughes, J. S., Lienhardt, H. F. 1936. Effects of low phosphorus rations on growing pigs. *J. Agric. Res.* 52:149–59
70. Meiter, A., Green, H. H. 1931. Aphosphorosis in ruminants. *Nutr. Rev.* 1:359
71. Day, H. G., McCollum, E. V. 1939. Mineral metabolism, growth, and symptomatology of rats on diet extremely deficient in phosphorus. *J. Biol. Chem.* 130:269–83
72. Freeman, S., McLean, F. C. 1941. Experimental rickets; blood and tissue changes in puppies receiving diet very low in phosphorus with and without vitamin D. *Arch. Pathol.* 32:387–408
73. Junos, J. H. 1934. The influence of the removal of the parathyroid glands on the development of rickets in rats. *J. Biol. Chem.* 106:701–5
74. Roof, B. S., Piel, C. F., Gordan, G. S. 1972. Nature of defect responsible for familial vitamin D-resistant rickets (VDRR) based on radioimmunoassay for parathyroid hormone (PTH). *Trans. Assoc. Am. Physicians* 85:172
75. Ben-Isaac, C., Massry, S. G., Rosenfeld, S., Kleeman, C. R., Bick, M. 1974. Evidence for humoral factor responsible for the hypercalciuria of phosphate depletion. *J. Clin. Invest.* 53:5a
76. Territo, M. C., Tanaka, K. R. 1974. Hypophosphatemia in chronic alcoholism. *Arch. Intern. Med.* 134:445–47

Ann. Rev. Med. 1978. 29:191–203
Copyright © 1978 by Annual Reviews Inc. All rights reserved

IMMUNOLOGICAL ASPECTS OF RENAL TUBULAR AND INTERSTITIAL DISEASES[1]

◆7276

Robert T. McCluskey, M.D. and Robert B. Colvin, M.D.
Department of Pathology, Massachusetts General Hospital, Boston,
Massachusetts 02114

Considerable effort over the last 25 years has been directed at the elucidation of the immunological basis of glomerular diseases. As a result, two major forms of antibody-mediated glomerular injury have been recognized; that due to immune complexes and that due to antibodies reactive with the glomerular basement membrane. Only recently have investigations focussed on immunologic mechanisms in tubulointerstitial nephritis, which comprises a large, diverse group of clinically important renal diseases. Already it is clear that analogous immunological mechanisms may cause tubulointerstitial damage.

Two types of antibody-mediated tubulointerstitial nephritis have been described in experimental animals (1, 2). One is mediated by autoantibodies reactive with the tubular basement membrane (anti-TBM disease) and is characterized by continuous (linear) accumulation of immunoglobulin along the TBM, as seen by immunofluorescence. The second is mediated by accumulation of immune complexes along the TBM or in the interstitium, and is manifested by granular deposits of immunoglobulins and complement. Evidence has been obtained, principally through immunofluorescence, that these mechanisms cause renal disease in man. Experimental studies also indicate that cell-mediated reactions can occur in the renal interstitium; however, the importance of cell-mediated sensitivity in human renal disease is uncertain, except in allografts. We review here experimental models of immunologically mediated tubulointerstitial disease and what appear to be corresponding lesions in man. In addition, drug-induced interstitial nephritis (for which there is no model) is discussed.

---

[1] Supported in part by a National Institutes of Health Grant 1 R01 AM 18729-01.

0066-4219/78/0401-0191$01.00

TUBULOINTERSTITIAL IMMUNE-COMPLEX LESIONS

Renal tubulointerstitial immune-complex deposits have been described in rabbits or rats sensitized with homologous renal tissue in adjuvant, and in rabbits with renal allografts (3–5). After several weeks, rabbits repeatedly injected with a soluble extract of renal cortex in adjuvant developed tubulointerstitial nephritis, with interstitial fibrosis, interstitial mononuclear cell infiltration, and tubular damage (5). The glomeruli were generally normal. Granular deposits of IgG and C3 were seen along the TBM of proximal tubules. After many months the deposits were sometimes no longer present, although tubulointerstitial damage remained. Anti-TBM antibodies were not demonstrable in the serum or in eluates of renal tissue, but there were antibodies directed against antigens in the cytoplasm of proximal tubular cells. We therefore suggested that the immune complexes were formed in situ, as a result of antigens exiting from tubular cells and meeting autoantibodies at the level of the TBM.

Tubulointerstitial immune-complex deposits also may form in situ in rats immunized with Tamm-Horsfall protein, a glycoprotein found in the distal nephron and a major constituent of casts (6). The deposits were found along the ascending thick limb of the loops of Henle, the macula densa, and part of the distal tubule; these are believed to be the sites of Tamm-Horsfall production. Mild interstitial mononuclear infiltration occurred in the vicinity of the deposits. Rats immunized with homologous kidney tissue in Freund's adjuvant sometimes developed broad deposits of IgG along the basal portions of Henle's loops and distal tubules (7), but the antigen was not identified. Local formation of complexes could probably occur only if the antigens are released into the extracellular space. Most tissue-specific autoantibodies directed against intracellular antigens do not have access to their corresponding antigen, and do not therefore cause local immune-complex formation.

Recently Brentjens et al (8), in a study of chronic serum sickness in rabbits induced by multiple daily injections of bovine serum albumin (BSA), found that immune-complex deposition occurs in various extraglomerular sites as well as in glomeruli. Immunofluorescence showed deposits of BSA, IgG, and IgM along the TBM, in Bowman's capsule, and around peritubular capillaries; electron microscopy showed dense deposits in these locations. Interstitial fibrosis, accumulation of mononuclear cells and neutrophils, as well as tubular cell damage, were present. The findings in this model indicate that the extraglomerular deposition occurs because very large amounts of circulating complexes are formed, which suggests that saturation of the usual removal mechanisms is involved. It is also possible, as noted above, that the complexes are formed locally, or that special receptors are involved in their localization.

In most models with tubulointerstitial immune deposits, interstitial inflammation and fibrosis occur, as well as tubular cell damage. In the study of Brentjens et al (8) a positive correlation was found between deposits and tubulointerstitial damage, which supports a pathogenic role of the complexes. No direct evidence exists concerning the secondary mechanisms by which the complexes mediate the injury. Moreover, even though complexes are present, other pathogenetic mechanisms (immunologic or nonimmunologic) may also contribute to the damage. A role for the infiltrating cells has been suggested in NZB/W mice, which develop severe interstitial nephritis in addition to glomerular lesions (9, 10). Although immunoglobulin deposition in glomeruli correlates with glomerular damage, a similar association is not found between the extent of tubulointersitital deposits and interstitial inflammation (9, 10). The tubular cell damage occurs near areas of mononuclear cell accumulation, which suggests that the infiltrating cells contribute to the cell injury (10). However, it is possible even here that immune complexes initiate the damage, but are present only intermittently or in undetectable amounts.

Following observations in experimental models, tubulointerstitial immunoglobulin and complement deposits were identified in man (11–13). The deposits had apparently previously been overlooked, or perhaps had been misinterpreted as reabsorption droplets, or as proteins present in peritubular capillaries. Even now, there may occasionally be such difficulties in interpretation. Electron microscopic studies may help by revealing dense deposits along or within tubular and capillary basement membranes or in the interstitium; however, because of the focal nature of the deposits, this is of limited usefulness.

The condition in which tubulointerstitial deposits are most often found is lupus nephritis (11–13). The deposits are often quite focal, but may be very widespread. By immunofluorescence, they appear to be located either along or within TBMs, around peritubular capillaries, or free in the interstitium; electron microscopic observations have substantiated these interpretations. The deposits invariably contain IgG and C3, and often IgM and/or IgA. In a few instances DNA has been demonstrated in the deposits, which provides additional evidence that they are immune complexes (13). In two reports, deposits were present in about half or more of the cases studied (12, 13). In general, interstitial inflammation and fibrosis and tubular cell damage appear to correlate with the abundance of the extraglomerular deposits. However, interstitial mononuclear cell infiltrates are sometimes found in the kidneys of patients with lupus nephritis with few if any tubulointerstitial deposits, similar to NZB/W mice. In most cases of lupus nephritis, the severity of the tubulointerstitial damage also parallels that of the glomerular disease. However, prominent tubulointerstitial nephritis has been observed with only mild glomerular involvement in a patient with renal failure (14, 15). Even when associated with glomerular lesions, the tubulointerstitial damage probably contributes to renal functional impairment.

Tubulointerstitial deposits have been described in the absence of glomerular disease in only a few other situations, namely in renal allografts (16), in a patient with Sjögren's syndrome (17), and in several patients with idiopathic interstitial nephritis (11). Tubulointerstitial deposits are sometimes much more abundant than glomerular deposits, which suggests that special factors are involved in their localization, rather than mere saturation of the sites in which complexes usually deposit.

Other forms of glomerulonephritis in which tubulointerstitial deposits have been described include glomerulonephritis with mixed cryoglobulinemia, membranoproliferative glomerulonephritis, and "rapidly progressive" (crescenteric) glomerulonephritis (11, 12). With the exception of mixed cryoglobulinemia, extraglomerular deposits are uncommon. In some cases of mixed cryoglobulinemia, widespread tubulointerstitial deposits are found, containing IgM, IgG, and C3. Rheumatoid factor activity has been found in the deposits (R. B. Colvin, unpublished), which is consistent with the interpretation that IgM–anti-IgG complexes are important.

Aside from systemic lupus erythematosus, where antigen (DNA) has been demonstrated in deposits, and in mixed cryoglobulinemia, where anti-IgG antibodies and IgG have been identified, there is no evidence concerning the identity of the antigen presumed to be present in the tubulointerstitial deposits, so that direct proof that they represent immune complexes is lacking (as is true in most glomerular diseases believed to be mediated by immune complexes).

ANTI–TUBULAR BASEMENT MEMBRANE ANTIBODY DISEASE

In 1971 Steblay & Rudofsky (18) reported that guinea pigs developed a fatal tubulointerstitial nephritis after immunization with the insoluble fraction of rabbit renal cortex, rich in tubules, in complete Freund's adjuvant. Light microscopy revealed interstitial mononuclear cells, multinucleated peritubular giant cells, and tubular cell damage. Immunofluorescence revealed linear deposition of IgG along most of the cortical TBMs. High titers of anti-TBM antibodies were detected in the sera by indirect immunofluorescence. Similar lesions occur in guinea pigs sensitized with bovine TBM (19, 20) and in certain strains of rats immunized with homologous or heterologous TBM (21, 22). In both species, transfer of serum containing anti-TBM antibodies to normal recipients results in renal lesions with the same immunofluorescence and histologic features as in actively immunized animals (20, 21, 23).

Precisely how the anti-TBM antibodies bring about tubular damage is not known. Guinea pigs genetically deficient in C4 are as susceptible to anti-TBM disease as normal animals (24). Although C3 is sometimes seen along the TBM, it may be absent, even in animals with severe disease (18–20). Because guinea pigs depleted of C3 by cobra venom factor are resistant to passively transferred disease, a pathogenic role of complement has been suggested (25), but cobra venom has other effects that might inhibit the development of renal lesions.

A recent study indicates that anti-TBM antibodies may amplify the disease by stimulating the production of further autoantibodies (C. L. Hall, R. B. Colvin, K. Carey, R. T. McCluskey, *J. Exp. Med.,* in press). This conclusion was reached in passive transfer studies in strain XIII guinea pigs. We began the investigation to determine whether IgG_1 or IgG_2 anti-TBM antibodies (or both) mediated the disease. Recipients of either IgG_1 or IgG_2 developed typical anti-TBM disease, as seen on day 14. This was surprising, in view of the known differences in the biological effects of the two isotypes. Analysis of the recipients' sera 14 days after they had received either IgG_1 or IgG_2 showed unexpectedly high levels of anti-TBM antibodies in both IgG_1 and IgG_2 fractions. In contrast, in recipients of similarly fractionated control IgG_1 and IgG_2 anti-bovine γ-globulin (BGG) antibodies, the titers declined markedly and anti-BGG activity was found only in the IgG fraction that had been transferred. These findings indicated that the recipients of antibodies were actually stimulated to produce their own anti-TBM antibodies. These experiments also showed that the autoantibodies produced by the recipient participated in the renal damage. It is not known whether the stimulation of autoantibodies by autoantibodies (which we call autoimmune amplification) is a common phenomenon; in most situations it would be difficult or impossible to detect. However, when it does occur, it may provide a mechanism for the intensification or perpetuation of autoimmune diseases.

A central question in anti-TBM disease is the pathogenic significance of the infiltrating mononuclear cells. In the guinea pig a majority of the infiltrating cells appear to be monocytes or macrophages, although numerous lymphocytes are also present. Macrophages have a wide assortment of digestive enzymes, and presumably have the potential to destroy the TBM, but whether they initiate the destruction of the TBM or merely ingest already-damaged membranes is not known. Electron microscopic observations suggest that the giant cells are formed by fusion of macrophages, which envelop fragments of TBM (21). However, giant cells are not an essential feature of anti-TBM disease in guinea pigs (19).

The nature of the infiltrating cells in anti-TBM disease has been studied by a rosetting technique that permits identification of B cells and monocytes in tissue sections (20). Many of the mononuclear cells bound IgG-coated red cells, which indicates they were monocytes or macrophages bearing Fc receptors (B cell Fc receptors do not bind in this assay). No C3 receptor–bearing lymphocytes were detected, which suggests that many of the lymphocytes were T cells or null cells. However, some B cells must have participated, because small numbers of plasma cells appeared later in the reaction. In lethally radiated guinea pigs reconstitution with normal bone marrow cells is essential for the development of disease after passive transfer of anti-TBM serum, which indicates that the bone marrow is a major source of the infiltrating cells (26). Since recipients of anti-TBM antibodies produce their own autoantibodies (see above), the protective effect of radiation may be attributable in part to suppression of antibody production rather than simply to elimination of cells that infiltrate the kidneys. Although the mononuclear nature of the infiltrate suggests a cell-mediated reaction, attempts to produce lesions by transfer of lymphoid cells have been unsuccessful (20, 27). Furthermore, lesions comparable to those in actively immunized animals can be produced by transfer of serum as early as three days, which indicates that antibodies alone are capable of initiating the process (5, 7). Nevertheless, it is conceivable that the combination of antibody with the TBM results in the formation of a conjugate that stimulates the development of delayed reactivity (as well as antibody production), and that this plays a role in the later stages of the disease.

Anti-TBM disease can be produced only in certain rat strains (21, 22). The resistant strains (Lewis, Wistar-Furth, and Maxx) lack the pertinent TBM antigens, as demonstrated by the failure of anti-TBM sera to bind their kidneys. This allotypic specificity is not determined by the major histocompatibility locus in the rat (Ag-B); Fisher 344 and Lewis are Ag-B identical, yet discordant in the expression of the TBM antigen. The same anti-TBM antibodies that fail to react with the resistant strains of rats do react with the TBM of a variety of other species.

Another recently appreciated aspect of anti-TBM disease is the genetic influence of the immune response to TBM antigen (28, 29). Strain XIII guinea pigs develop much more severe anti-TBM disease when immunized with rabbit TBM than strain II guinea pigs (28). Strain II animals produce little or no antibody reactive with guinea pig TBM, in contrast with strain XIII guinea pigs. The strain II kidneys possess the TBM antigen(s) that bind anti-TBM antibodies from strain XIII animals; thus, their failure to develop disease is not caused by the absence of the pertinent TBM antigen. Strain II animals also fail to develop disease even when given anti-TBM antibodies; this is probably due to a failure of the recipients to produce their own anti-TBM antibodies. Certain rat strains also are low responders to TBM (22). In the guinea pig a dominant immune-response (Ir) gene(s) linked to the major histocompatibility locus controls the autoimmune anti-TBM response (28).

Since the demonstration of anti-TBM disease in animals, small numbers of cases of human anti-TBM disease have been recognized. The evidence for anti-TBM antibodies consists of finding linear accumulation of IgG and C3 along cortical tubules, primarily proximal tubules, as well as demonstrating anti-TBM activity by indirect immunofluorescence in the sera and in eluates of kidneys. (Irregular linear deposits of C3 alone are sometimes seen in normal kidneys). In all instances there has been interstitial inflammation, often with fibrosis, and tubular cell damage. However, multinucleated giant cells have been reported in only one case (2).

A majority of patients with anti–glomerular basement membrane (anti-GBM) disease with Goodpasture's syndrome (or, less commonly, rapidly progressive glomerulonephritis) have associated anti-TBM disease (12). Moreover, eluates of renal tissue have exhibited anti-TBM activity, in addition to antibodies that react with glomerular and often pulmonary basement membranes. Absorption of eluates with solubilized GBM has been shown to remove the anti-TBM activity (30). These results indicate that the responsible antigenic determinants are present both in the TBM and GBM.

Anti-TBM antibodies have also been reported in several situations in the absence of anti-GBM antibodies, including immune-complex glomerulonephritis (31–33), drug-induced interstitial nephritis (see below) (34–36), idiopathic interstitial nephritis (37), and renal allografts (38, 39). The explanation for the association of immune-complex glomerular disease and anti-TBM antibodies (which has been observed in more instances than would appear likely on a chance basis) is not clear. Possibly the glomerular disease, which in at least one case clearly preceded the anti-TBM disease (33), brings about secondary tubular damage, which in turn triggers an autoimmune response. It may also be that damage to the GBM stimulates autoantibodies that are directed primarily against cross-reacting antigens in the TBM.

Allotypes of TBM antigens have been detected in man (39), and some human anti-TBM sera react with the Brown Norway (BN) but not the Lewis TBM (34, 37). The anti-TBM antibodies found in a patient with a renal allograft reacted with the TBM in the allograft and all other human kidneys tested, but not with the patient's own kidneys (38). However, another human renal allograft recipient developed anti-TBM antibodies that reacted with his own kidneys as well as the transplanted and other kidneys (39), which indicates that allografts may sometimes stimulate antibody production to autologous TBM antigens.

CELL-MEDIATED TUBULOINTERSTITIAL DISEASE

The possibility that delayed hypersensitivity directed against autologous antigens can produce interstitial nephritis is supported by the observation that some rats injected with homologous kidney preparations develop interstitial mononuclear cell infiltrates; the fact that these infiltrates often occur early, before evidence of autoantibody production, favors a cell-mediated mechanism (T. Sugisaki, J. Klassen, G. A. Andres, R. T. McCluskey, unpublished). In addition, Grupe (40) found that lymphocytes from rats immunized with kidney tissue produced macrophage-migration inhibitory factor in the presence of renal antigens. Furthermore, renal antigens also elicit delayed skin reactions and provoke lymphocyte blast transformation (41). However, early interstitial nephritis following immunization with kidney homogenates does not occur consistently and conclusive evidence that the lesions can be transferred with cells has not been obtained.

It has only recently been shown that delayed hypersensitivity can be elicited in renal tissue with exogenous antigens (42). Guinea pigs and rats were sensitized to BGG using methods that favor the development of delayed hypersensitivity. Direct injection of aggregated BGG into the renal cortex resulted in a mononuclear interstitial infiltrate, with focal tubular destruction. Reactivity could be transferred with lymph node cells, but not with serum. Although reactions were consistently produced with heat-aggregated BGG, soluble BGG generally failed to elicit reactions, presumably because it rapidly diffused from the injection site. In contrast, soluble BGG was fully as effective as aggregated material in eliciting delayed reactions in the skin. These findings show that the conditions required for the development of delayed reactions are not the same in all tissues, and suggest that in the kidney only those antigens that are aggregated, particulate (such as bacteria), or fixed (such as autologous antigens) are highly effective in eliciting delayed reactions.

Some effort has been made to relate cellular reactivity to bacterial antigens to the pathogenesis of pyelonephritis (43). In experimental models mononuclear cells appear very rapidly in the infiltrate. Furthermore, Cotran (43) showed that rats with pyelonephritis induced by streptococci developed cellular reactivity to the bacteria, as indicated by lymphocyte stimulation in vitro. In addition, bacterial antigens were shown by the immunofluorescence in amorphous and particulate form, both within leukocytes and free in the interstitium. Obviously, if bacterial antigens do elicit delayed reactions in the kidney, this would be only one aspect of the pathogenesis of pyelonephritis. Other mechanisms by which bacteria produce tissue damage might be of greater importance. Furthermore, the delayed reaction might be important in defense, as well as in tissue injury. It seems unlikely that delayed reactivity against bacterial antigens could provide a mechanism for perpetuation of pyelonephritis beyond the period when viable organisms are present in the kidney, since this would require persistence of bacterial antigen in sufficient amounts and in appropriate physical form. Indeed, in both human and experimental pyelonephritis, bacterial antigens are usually not demonstrable in kidneys after eradication of bacteria, and when present are found only within macrophages (43). There is no substantial evidence that bacterial infections of the kidney induce an autoimmune response against renal antigens (43).

In many forms of renal disease in man, with or without associated glomerular lesions, interstitial mononuclear cell infiltrates are conspicuous. To what extent these infiltrates represent delayed hypersensitivity reactions is unknown. This uncertainty stems largely from the lack of reliable criteria for the identification of cell-mediated reactions. Histologic features do not provide a reliable guide. Mononuclear cell infiltrates can be seen in certain antibody-mediated reactions (as in anti-TBM disease), as well as in nonimmunological reactions. Even the use of newer techniques that permit the identification of different types of mononuclear cells in tissue sections has not yet made it possible to identify delayed reactions, nor to assess the functional activity of the cells. The in vitro demonstration of cell-mediated reactivity against a particular antigen does not establish an in vivo role for this mechanism in the production of disease. The most reliable evidence for a pathogenic role of delayed sensitivity—transfer of disease with lymphocytes—obviously cannot be obtained in man. It is apparent that new approaches to the recognition of delayed reactions and for evaluation of their significance are needed. Nevertheless, there are reasons to suspect that cell-mediated reactions occur in human renal disease and that they may be directed against either autologous or exogenous antigens.

Rocklin, Lewis & David (44) demonstrated production of migration inhibition factor by lymphocytes from some patients with anti-GBM disease when their lymphocytes were incubated with preparations of GBM (which undoubtedly included a variety of renal antigens). Others made similar observations in patients with focal or diffuse proliferative glomerular disease, as well as in some patients with renal vascular disease (45). Delayed hypersensitivity to bacterial antigens may play a role in the pathogenesis of pyelonephritis (see above). Some forms of drug-induced interstitial nephritis may be due, at least in part, to cell-mediated mechanisms, as discussed below.

DRUG-INDUCED TUBULOINTERSTITIAL NEPHRITIS

A variety of drugs are known to induce acute tubulointerstitial nephritis in man, characterized histologically by interstitial edema, tubular cell damage, and irregular interstitial leucocytic infiltration. The infiltrate usually consists principally of mononuclear phagocytes and lymphocytes, but eosinophils are generally present, often in large numbers, and plasma cells and small numbers of basophils are sometimes also found. Tubules are invaded by mononuclear cells and eosinophils. The blood vessels and glomeruli appear normal. Clinically the renal disease is manifested by hematuria, proteinuria, enlargement of the kidneys, and in severe cases acute renal failure. Mild cases may go unrecognized. The drugs most frequently incriminated are methicillin, ampicillin, rifampicin, and phenindione [for complete list, see (46)]. With certain agents, especially methicillin and related antibiotics, there is considerable evidence that the lesions result from immunological mechanisms, rather than from toxic effects (46–48). Thus, only a small proportion of patients receiving the antigen develop interstitial nephritis, and this is not dose related. Moreover, most of the patients show evidence of hypersensitivity, with fever, rash, arthralgia, and eosinophilia. Repeated exposure to the same drug, or a closely related agent, sometimes brings about recurrence of the interstitial nephritis.

Certain infections, notably scarlet fever and diphtheria, are sometimes complicated by acute interstitial nephritis (presumably not due to bacterial invasion of the kidney). The possibility that the infection caused the interstitial nephritis seen in some patients receiving antibiotic treatment is therefore worth considering. However, the same renal lesions have been observed in patients treated for a variety of infections, most of which have not in the past been associated with interstitial nephritis; moreover, severe interstitial nephritis has been observed to develop in some patients given methicillin prophylactically (35).

In an individual patient it is often difficult to determine whether renal damage is due to toxic effect of the drug, an immunological reaction initiated by the drug, an infectious process, or some other unknown cause. Moreover, in patients receiving multiple drugs, it may be difficult to identify the offending agent. Identification of the cause of renal damage is usually easier when renal tissue is available for pathological studies, but even these studies are not always conclusive. In instances in which hypersensitivity is responsible, there is usually conspicuous interstitial leucocytic infiltration and edema, features that are generally lacking or mild in toxic injury. There are probably no criteria for identification of acute interstitial nephritis associated with infection. In addition, especially with small samples, it is sometimes uncertain whether one is dealing with interstitial nephritis at all, rather than simply an unimportant interstitial infiltrate, as may be seen around areas of ischemia or other forms of damage.

The immunological mechanisms usually responsible for drug-induced interstitial nephritis have not been elucidated. Probably several mechanisms are involved and vary in importance from one case to another. In some cases of methicillin- or penicillin-associated interstitial nephritis, a derivative of the antibiotic is bound to certain sites in renal tissue, in particular to the TBM (34, 47, 48). Assuming that renal fixation is important, two general possibilities can be considered: (*a*) The drug hapten binds to renal tissues in all patients receiving the drug, and the development of lesions depends on an unusual immune response (either humoral or cell-mediated); or (*b*) The binding of hapten occurs only in those patients who develop interstitial nephritis. There is conflicting evidence concerning this point. Border et al (34) were unable to find methicillin antigen in the kidneys of animals given methicillin; however, they did not provide information concerning the dosage or time of administration. In contrast, Colvin et al (48) described binding of antipenicillin antibodies to the TBM and interstitium in autopsy specimens of histologically normal kidneys from patients who received large amounts of penicillin shortly before death.

Several observations indicate that cell-mediated reactivity (presumably directed against a conjugate of the drug and a structural renal protein) may be important: the predominance of mononuclear cells in the infiltrate, the scarcity or absence of immunoglobulins and C3 deposits in the kidneys in many cases, and the presence of cellular reactivity to drug. However, as noted above, none of these findings proves that the lesions are due to delayed sensitivity.

The presence of eosinophils points to an antibody-mediated component. Increased levels of IgE in the serum (49), IgE-containing plasma cells (in a case of interstitial nephritis associated with phenobarbital) (50), and the presence of basophils (48) all suggest a role for reaginic antibodies in some cases.

Anti-TBM antibodies were present in a case of drug-induced interstitial nephritis reported by Border et al (34). Linear accumulation of IgG and C3 was found along the TBM, together with a methicillin-derived antigen. Circulating anti-TBM antibodies were found. It was suggested that dimethoxyphenyl penicillin, which is largely secreted by the proximal tubules, binds to the TBM and results in the formation of a methicillin-TBM conjugate that stimulates anti-TBM autoantibody production. It is also possible, however, that anti-TBM antibodies result from tubular damage produced in some other way. In any case, anti-TBM antibodies have been reported in only a few cases and would appear to be an uncommon and secondary pathogenetic mechanism.

CONCLUSION

Four general conclusions can be stated. First, renal tubulointerstitial immune-complex deposition has been described in several models, and is associated with tubulointerstitial damage. Similar deposits have been described in man. They appear to be common in lupus nephritis, probably in the renal disease associated with mixed cryoglobulinemia, and possibly in renal allografts. The role of extraglomerular deposits in Sjögren's syndrome deserves further study, since otherwise unexplained interstitial nephritis is often found in these patients. It seems unlikely that future studies will reveal tubulointerstitial deposits in an appreciable percentage of cases of most forms of glomerulonephritis or in idiopathic interstitial nephritis. However, the remote possibility remains that immune complexes do account for certain forms of unexplained tubulointerstitial damage, even though they are not detectable by immunofluorescence. Second, anti-TBM antibodies cause severe tubulointerstitial nephritis in animals. Anti-TBM antibodies have been found in man in small numbers of patients, most often in association with anti-GBM diseases. In view of the resemblance of the findings in man to the experimental models, it seems probable that the anti-TBM antibodies are of pathogenetic significance. Third, delayed hypersensitivity reactions can occur in the kidneys of animals. It is quite likely that the interstitial mononuclear cell infiltrates seen in some human renal diseases represent cell-mediated reactions; however, proof will require the development of methods to identify such reactions in vivo. Fourth, a common form of acute interstitial nephritis is believed to be caused by hypersensitivity to a variety of drugs. Cellular reactivity, IgE, and rarely anti-TBM antibodies may participate in this form of renal injury.

Literature Cited

1. McCluskey, R. T., Klassen, J. 1973. Immunologically mediated glomerular, tubular and interstitial renal disease. *N. Engl. J. Med.* 288:564–70
2. Andres, G. A., McCluskey, R. T. 1975. Tubular and interstitial renal disease due to immunologic mechanisms. *Kidney Int.* 7:271–89
3. Unanue, E. R., Dixon, F. J., Feldman, J. D. 1967. Experimental allergic glomerulonephritis induced in the rabbit with homologous renal antigens. *J. Exp. Med.* 125:163–75
4. Klassen, J., Milgrom, F. 1969. Autoimmune concomitants of renal allografts. *Transplant. Proc.* 1:605–8
5. Klassen, J., McCluskey, R. T., Milgrom, F. 1971. Nonglomerular renal disease produced in rabbits by immunization with homologous kidney. *Am. J. Pathol.* 63:333–50
6. Hoyer, J. R. 1976. Autoimmune tubulointerstitial nephritis induced in rats by immunization with rat Tamm-Horsfall (TH) urinary glycoprotein. *Kidney Int.* 10:544 (Abstr.)
7. Klassen, J., Suqisaki, T., Milgrom, F., McCluskey, R. T. 1971. Studies on multiple renal lesions in Heymann nephritis. *Lab. Invest.* 25:577–85
8. Brentjens, J. R., O'Donnell, D. W., Pawlowski, I. B., Andres, G. A. 1974. Extra-glomerular lesions associated with deposition of circulating antigen-antibody complexes in kidneys of rabbits with chronic serum sickness. *Clin. Immunol. Immunopathol.* 3:112–22
9. Hurd, E. R., Ziff, M. 1977. Quantitative studies of immunoglobulin deposition in the kidney, glomerular cell proliferation and glomerulosclerosis in NZB/NZW F. hybrid mice. *Clin. Exp. Immunol.* 27:261–68
10. Hurd, E. R., Ziff, M. 1977. Influence of interstitial infiltration on tubule cells in kidneys of NZB/NZW FI (B/W) hybrid mice and the effect of cyclophosphamide. *Arthritis Rheum.* 20:122 (Abstr.)
11. Klassen, J., Andres, G. A., Brennan, J. C., McCluskey, R. T. 1972. An immunologic renal tubular lesion in man. *Clin. Immunol. Immunopathol.* 1:69–83
12. Lehman, D. H., Wilson, C. B., Dixon, F. J. 1975. Extraglomerular immunoglobulin deposits in human nephritis. *Am. J. Med.* 58:765–86
13. Brentjens, J. R., Sepulveda, M., Baliah, T., Bentzel, C., Erlanger, B. F., Elwood, C., Montes, M., Hsu, K. C., Andres, G. A. 1975. Interstitial immune complex neph-

ritis in patients with systemic lupus erythematosus. *Kidney Int.* 7:342–50

14. Case Records of the Massachusetts General Hospital. Case 2-1976. *N. Engl. J. Med.* 294:100–5
15. Nicastri, A. D., Chen, C. K., Rao, T. K. S., Ginzler, E. M., Kaplan, D., Friedman, E. A. 1975. Renal disease with tubular immunofluorescence deposits. *Kidney Int.* 8:452 (Abstr.)
16. Andres, G. A., Accinni, L., Hsu, K. C., Penn, I., Porter, K. A., Randall, J. M., Seegal, B. C., Starzl, T. E. 1970. Human renal transplants: III. Immunopathologic studies. *Lab. Invest.* 22:588–604
17. Sawhney, A. S., Winer, R. L., Cohen, A. H., Gorman, J. T. 1975. Sjogren's syndrome with immune complex tubulointerstitial renal disease. *Kidney Int.* 6:453 (Abstr.)
18. Steblay, R. W., Rudofsky, U. 1971. Renal tubular disease and autoantibodies against tubular basement membrane induced in guinea pigs. *J. Immunol.* 107:589–94
19. Lehman, D. H., Marquardt, H., Wilson, C. B., Dixon, F. J. 1974. Specificity of autoantibodies to tubular and glomerular basement membranes induced in guinea pigs. *J. Immunol.* 112:241–48
20. Van Zwieten, M. J., Bhan, A. K., McCluskey, R. T., Collins, A. B. 1976. Studies on the pathogenesis of experimental anti-tubular basement membrane nephritis in the guinea pig. *Am. J. Pathol.* 83:531–46
21. Sugisaki, T., Klassen, J., Milgrom, F., Andres, G. A., McCluskey, R. T. 1973. Immunopathologic study of an autoimmune tubular and interstitial renal disease in Brown Norway rats. *Lab. Invest.* 28:658–71
22. Lehman, D. H., Wilson, C. B., Dixon, F. J. 1974. Interstitial nephritis in rats immunized with heterologous tubular basement membrane. *Kidney Int.* 5:187–95
23. Steblay, R. W., Rudofsky, U. 1973. Transfer of experimental autoimmune renal cortical tubular and interstitial disease in guinea pigs by serum. *Science* 180:966–68
24. Rudofsky, U. H., McMaster, P. R. B., Ma, W., Steblay, R. W., Pollara, B. 1974. Experimental autoimmune renal cortical tubulointerstitial disease in guinea pigs lacking the fourth component of complement (C4). *J. Immunol.* 112:1387–93
25. Rudofsky, U. H., Steblay, R. W., Pollara, B. 1975. Inhibition of experimental autoimmune renal tubulointerstitial disease in guinea pigs by depletion of complement with cobra venom factor. *Clin. Immunol. Immunopathol.* 3:396–402
26. Rudofsky, U. H., Pollara, B. 1976. Studies on the pathogenesis of experimental autoimmune renal tubulointerstitial disease in guinea pigs. II. Passive transfer of renal lesions by anti-tubular basement membrane autoantibody and nonimmune bone marrow cells to leukocyte-depleted recipients. *Clin. Immunol. Immunopathol.* 6:107–14
27. Lehman, D. H., Wilson, C. B. 1976. Role of sensitized cells in antitubular basement membrane interstitial nephritis. *Int. Arch. Allergy Appl. Immunol.* 51:168–74
28. Hyman, L. H., Steinberg, A. D., Colvin, R. B., Bernard, E. F. 1977. Immunopathogenesis of autoimmune tubulointerstitial nephritis. II. Role of an immune response gene linked to the major histocompatibility locus. *J. Immunol.* 117:1894–97
29. Hyman, L. R., Colvin, R. B. Steinberg, A. D. 1976. Immunopathogenesis of autoimmune tubulointerstitial nephritis. I. Demonstration of differential susceptibility in strain II and strain XIII guinea pigs. *J. Immunol.* 116:327–35
30. McPhaul, J. J., Dixon, F. J. 1970. Characterization of human anti–glomerular basement membrane antibodies eluated from glomerulonephritic kidneys. *J. Clin. Invest.* 49:308–17
31. Levy, M., Gagnadoux, M. F., Habib, R. 1974. An immunologic Fanconi syndrome. *Int. Symp. Pediatr. Nephrol., 3rd,* Washington, DC
32. Harner, M. H., Nolte, M., Wilson, C. B., Talwalkar, Y. B., Musgrave, J. E., Brooks, R. E., Campbell, R. A. 1974. Anti-TBM antibody and nephrotic syndrome associated with milk allergy. See Ref. 31
33. Morel-Maroger, L., Kourilsky, O., Mignon, F., Fichet, G. 1974. Anti–tubular basement membrane antibodies in rapidly progressive post-

streptococcal glomerulonephritis: Report of a case. *Clin. Immunol. Immunopathol.* 2:185–94

34. Border, W. A., Lehman, D. H., Egan, J. D., Sass, H. J., Glade, J. E., Wilson, C. B. 1974. Antitubular basement membrane antibodies in methicillin associated interstitial nephritis. *N. Engl. J. Med.* 291:381–84
35. Olsen, S., Asklund, M. 1976. Interstitial nephritis with acute renal failure following cardiac surgery and treatment with methicillin. *Acta Med. Scand.* 199:305–10
36. Hyman, L. R., Ballow, M., Knieser, M. R. 1975. Diphenylhydantoin nephropathy: Evidence for an autoimmune pathogenesis. *Kidney Int.* 8:450 (Abstr.)
37. Bergstein, J. M., Litman, N. 1975. Interstitial nephritis with antitubular basement membrane antibody. *N. Engl. J. Med.* 292:875–78
38. Wilson, C. B., Lehman, D. H., McCoy, R. C., Gunnelli, C. J., Stickel, D. L. 1974. Antibular basement membranes after renal transplantation. *Transplantation* 18:447–52
39. Klassen, J., Kano, K., Milgrom, F., Menno, A. B., Anthone, S., Anthone, R., Sepulveda, M., Elwood, C. M., Andres, G. A. 1973. Tubular lesions produced by autoantibodies to tubular basement membrane in human renal allografts. *Int. Arch. Allergy Appl. Immunol.* 45:674–89
40. Grupe, W. E. 1968. An "in vitro" demonstration of cellular sensitivity in experimental autoimmune nephrosis in rats. *Proc. Soc. Exp. Biol. Med.* 127:1217–22
41. Litwin, A., Adams, L. E., Levy, R., Cline, S., Hess, E. B. 1971. Cellular immunity in experimental glomerulonephritis of rats: I. Delayed hypersensitivity and lymphocyte stimulation studies with renal tubular antigens. *Immunology* 20:755–66
42. Van Zweiten, M. J., Leber, P. D., Bhan, A. K., McCluskey, R. T. 1977. Experimental cell mediated interstitial nephritis induced with exogenous antigens. *J. Immunol.* 118:589–93
43. Cotran, R. S., Piessens, W. F. 1976. Pathogenesis of chronic pyelonephritis. *Proc. Int. Congr. Nephrol., 6th,* pp. 509–23. Basel: Karger
44. Rocklin, R. E., Lewis, E. J., David, J. R. 1970. *In vivo* evidence for cellular hypersensitivity to glomerular basement membrane antigens in human glomerulonephritis. *N. Engl. J. Med.* 283:497–501
45. Mahieu, P., Dardenne, M., Bach, J. F. 1972. Detection of humoral and cell-mediated immunity to kidney basement membranes in human renal diseases. *Am. J. Med.* 53:185–92
46. Mery, J. P., Morel-Moroger, L. 1976. Acute interstitial nephritis. A hypersensitivity reaction to drugs. *Proc. Int. Congr. Nephrol., 6th,* pp. 524–29. Basel: Karger
47. Baldwin, D. S., Levine, B. B., McCluskey, R. T., Gallo, G. R. 1968. Renal failure and interstitial nephritis due to penicillin and methicillin. *N. Engl. J. Med.* 279:1245–52
48. Colvin, R. B., Burton, N. E., Hyslop, N. E. Jr., Spitz, L., Lichtenstein, N. S. 1974. Penicillin-associated interstitial nephritis. *Ann. Int. Med.* 81:404–5
49. Ooi, B. S., First, M. R., Pesce, A. J., Pollak, V. E., Bernstein, I. L., Jao, W. 1974. IgE levels in interstitial nephritis. *Lancet* 1:1254–56
50. Faarup, P., Christensen, E. 1974. IgE containing plasma cells in acute tubulointerstitial nephropathy. *Lancet* 2:718

Ann. Rev. Med. 1978. 29:205–18
Copyright © 1978 by Annual Reviews Inc. All rights reserved

THE DISPOSITION AND EFFECTS OF SEDATIVES AND ANALGESICS IN LIVER DISEASE[1]

◆7277

Anastacio M. Hoyumpa, Jr., M.D.,
Robert A. Branch, M.D., M.R.C.P., and Steven Schenker, M.D.
Departments of Medicine and Pharmacology, Vanderbilt University School of Medicine and Veterans Administration Hospital, Nashville, Tennessee 37323

INTRODUCTION

The aim of drug therapy is to deliver a drug concentration sufficient to affect the site of action. Whether the desired intensity and duration of action are attained or exceeded depends largely on the efficiency of drug disposition and on the sensitivity of the receptor site(s). The pharmacologic effects of sedatives and analgesics have been shown to be exaggerated in some patients with significant liver disease or portasystemic shunts, with coma as a potential outcome (1). In recent years a number of studies revealed impaired disposition and altered response to such widely used drugs as tranquilizers, sedatives, and analgesics in the presence of significant hepatic dysfunction (2). The purpose of the present discussion is to review the disposition and effects of drugs in hepatic dysfunction, and thus provide a rational approach to the management of liver disease.

ROUTES OF HEPATIC DRUG METABOLISM

Of the many tissues that metabolize drugs, the liver is quantitatively the most active. Two major groups of chemical reactions are involved in the hepatic biotransformation of drugs (3). The first group involves oxidation, reduction, or hydrolysis. Of these, oxidation is the most common, and it is accomplished by the mixed function monooxygenase system located in the smooth endoplasmic reticulum of the hepatocyte. This enzyme system, the best studied component

---

[1] Supported by Grants 5R01-AA-00267-06 and GM 15431

0066-4219/78/0401-0205$01.00

of which involves cytochromes P450 and P448, supplies oxygen to form high energy complexes (4) that possess a marked affinity for a wide range of compounds. Having produced an alteration of the substrate, which renders the parent drug more polar and thus easier to eliminate, the monooxygenase system is released in an altered state, ready again to be formed into another high-energy complex. The transformation of the parent drug has variable pharmacologic effects, which may result in activation, alteration in pharmacologic effect, or inactivation of the agent. For example, diazepam (Valium®) and chlordiazepoxide (Librium®) are both metabolized via these pathways to pharmacologically active compounds (5–9). On the other hand, imipramine, which has phenothiazine-like properties, is converted to desmethylimipramine, an antidepressant (10), while many of the barbiturates are metabolized directly to inactive compounds (11).

The second major metabolizing enzyme system involves conjugation that couples the drug (or its metabolite) with an endogenous substance. For example, glucuronic acid, sulfate, or glutathione moeities may replace the hydrogen in OH, COOH, NH_2 or SH_2 groups. Conjugation usually abolishes the pharmacologic or toxic properties of the drug by making it more soluble and thus easier to excrete into urine or bile. Examples of such drugs are the benzodiazepines, oxazepam (12) and lorazepam (13), which are conjugated into pharmacologically inactive glucuronides and excreted primarily via the kidneys. Similarly the major route of morphine metabolism is by glucuronidation (14).

In the metabolism of drugs a number of reactions are usually involved either sequentially or concurrently. Moreover, there is as yet no substantial evidence to suggest that a disease process influences the chemical pathway along which an individual drug is metabolized in comparison to normal subjects. However, since many drugs are metabolized by more than one pathway, a disease process could conceivably influence which route is used, and this could affect the pharmacologic or toxic response and the overall efficiency of drug elimination.

FACTORS THAT MAY INFLUENCE DRUG ELIMINATION

Drug elimination is a complex process, and it may be affected by a number of factors, in addition to the presence of liver disease. These include the genetic phenotype of the individual (15), age (16), sex (17, 18), nutrition (19), diet (20–22), use of other therapeutic agents (23) or alcohol, and, for some drugs, renal function (24) and urinary pH (18). The important role of genetic structure in drug elimination was well validated by studies that showed much less individual variation in drug disposition between identical twins than between fraternal twins or unrelated persons, provided other variables are excluded (15). As the age of experimental animals increased, cytochrome P450 content and metabolizing-enzyme activity was reduced, and response to enzyme induction decreased (16). In man too, aging causes decreased elimination of some drugs (i.e. antipyrine (16)) and alters other aspects of drug disposition, such as volume of distribution (i.e. diazepam (25)) and plasma protein binding (i.e. antipyrine (16)). Some drugs are influenced by dietary intake not only in their absorption but also in their elimination. Thus antipyrine and theophylline are eliminated more rapidly following the intake of a high-protein, low-carbohydrate diet (20, 21). In view of the critical role of the mixed-function oxidase system in drug biotransformation, it is understandable that other drugs, by nonspecifically stimulating or competing for binding to the oxidase system, may alter the elimination of any given therapeutic agent. Finally, some drugs are eliminated in part by the kidneys, and others are importantly affected by the urinary pH (i.e. phenobarbital (26), methadone (18), and meperidine (27, 28)). It is evident therefore that drug elimination is governed by a multitude of factors, which accounts for the often wide variation in drug elimination among normal individuals. Lack of proper control of these variables in the past probably explains divergent reports on the effects of hepatic disease on the pharmacokinetics of various drugs.

PHARMACOKINETIC CONSIDERATIONS

A brief consideration of drug disposition and elimination (pharmacokinetics) is essential to assure an understanding of the terms used and to provide a conceptual framework for discussion of the mechanisms that govern drug elimination in health and in liver disease. The characterization of drug disposition has been approached in two ways.

1. The simplest approach to pharmacokinetics is the compartmental approach. This approach assumes that the drug is distributed in the body in a single compartment, the volume of which can be determined by dividing the dose given by the estimated plasma concentration prior to drug elimination. This parameter is referred to as the volume of distribution (V_d). This term is hypothetical and applies to the volume required for an equal distribution of the drug. However, it has no real anatomic or physiologic reality. On the other hand, the efficiency of the drug removal process is termed clearance (C) which is a measure of the rate at which the parent drug is removed from plasma or blood. The elimination of the drug from the body is conventionally expressed in terms of elimination half-life ($T_{\frac{1}{2}}$), defined as the time required to remove 50% of the initial drug concentration. Thus, $T_{\frac{1}{2}} = (0.693 \times V_d)/C$. It is evident that $T_{\frac{1}{2}}$ is dependent on both distribution and clearance, and is not solely a measure of drug elimination. It has become apparent that the pharmacokinetics or disposition of some drugs is more complex than can be explained by a single compartmental approach, and the concept of drug distribution between multiple compartments has been applied.

2. The physiologic approach to pharmacokinetics has significantly advanced our understanding of how various diseases influence drug disposition (29–32). This approach emphasizes that clearance is not only limited by drug-metabolizing activity, but also by factors that influence delivery of drug (hepatic blood flow) to the metabolizing enzymes. It characterizes hepatic clearance in terms of the basic rate-limiting parameters. These are: (*a*) the intrinsic free-drug clearance (C_{int}), which is a measure of drug-metabolizing activity when flow is not rate limiting; (*b*) liver blood flow (Q); (*c*) free fraction of drug in blood (f_B); (*d*) the presence or absence of portasystemic shunting of orally administered

drugs. Thus, $C = (Q \times C_{\text{int}} f_B/(Q + C_{\text{int}} f_B)$ This equation defines the relation between basic rate-limiting factors that determine hepatic clearance, and it also determines the steady-state blood concentration, hepatic extraction, the proportion of drug administered orally that reaches the systemic circulation, and bioavailability.

This important concept offers a new approach to the classification of drugs, and allows the physician to anticipate to what extent and duration a drug response might be influenced by the kinetic interaction in disease states. The classification is based on the intrinsic free clearance and protein-binding characteristics of drugs (Table 1). The differences between the groups are illustrated in Figure 1.

According to this concept, when a drug with high intrinsic clearance (group I, Table 1) is given by mouth and absorbed from the gastrointestinal tract, it then passes into the portal circulation. If the portal circulation is normal (no formation of collateral vessels), the drug passes to the liver where a large proportion of the drug is extracted, which results in a large "first-pass" effect equal to the hepatic extraction ratio. This then results in reduced bioavailability of the drug to the systemic circulation. If on the other hand, significant portasystemic shunting is present, then, although the free intrinsic clearance remains unchanged, a greater proportion of the drug reaches the systemic circulation, i.e. small first-pass effect. Variation of protein binding of drugs with high intrinsic clearance does not restrict elimination unless binding is exceptionally high, but changes in protein binding affect V_d and, more significantly, the steady-state concentration of the free drug at the receptor site. At present there are no known examples of drugs with high intrinsic free clearance and low protein binding.

In drugs with low free intrinsic clearance and low protein binding (group II), systemic clearance is low and independent of hepatic blood flow. In addition, organ extraction is low and changes with blood flow, so that the product of flow and extraction remains constant. Thus the major determinant is a change in free intrinsic clearance.

In contrast, if free intrinsic clearance is low but protein binding is high (group

Table 1 Classification of psychotropic drugs according to clearance and binding characteristics

Group	Free intrinsic clearance[a]	Protein binding[a]	Generic drug group
I	High	High	Morphine Meperidine Tricyclic antidepressant Phenothiazines
II	Low	Low	Barbiturates Nonbarbiturate hypnotics
III	Low	High	Benzodiazepine

[a] Drugs with hepatic extraction ratio of > 30% are artibrarily considered drugs with high intrinsic clearance, and protein binding > 85% is considered high value.

III), then protein binding allows less free (unbound) drug to be available for elimination. Thus, with very high protein binding, hepatic clearance becomes proportional to the presence of free drug. Changes in both intrinsic free clearance and binding are major determinants in drug disposition kinetics. Drugs with a low intrinsic free clearance have a small first-pass effect, and the development of portasystemic shunts does not influence bioavailability.

The approach presented here is an attempt to relate pharmacokinetic changes in liver disease to situations of clinical relevance. It may also provide a framework for anticipating the effects of liver disease on the disposition of drugs still to be studied.

EFFECT OF LIVER DISEASE ON THE DISPOSITION OF SEDATIVES AND ANALGESICS

Improved understanding of the physiology of drug handling allows us now to anticipate and rationalize changes in drug disposition in patients with liver disease. So far, however, only a limited number of studies in this area have appeared (mainly in patients with cirrhosis and acute hepatitis), and in some of these, avoidance of other drugs and alcohol was not strictly controlled, plasma protein drug binding was not measured, and the data were expressed only in terms of $T_{\frac{1}{2}}$, which limits adequate interpretation of the data. The available pharmacokinetic data on sedatives and analgesics are given in Table 2, and each group of drugs is discussed below.

Tranquilizers

BENZODIAZEPINES Benzodiazepines have attained widespread popularity as tranquilizers and sedatives. In man the benzodiazepines have low free intrinsic clearance and are highly protein bound. Thus clearance and bioavailability are independent of liver blood flow.

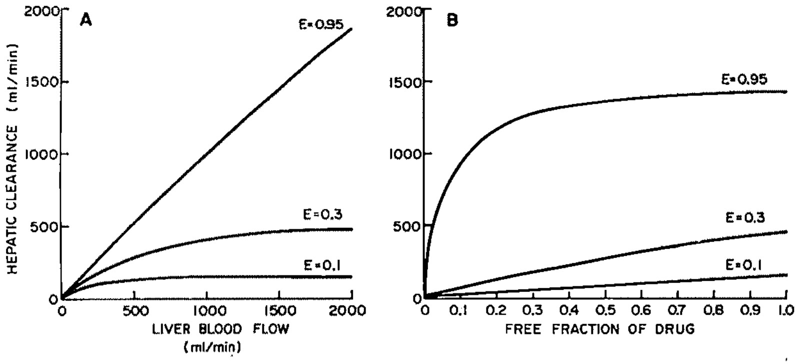


Figure 1 Theoretical relationship between hepatic clearance and liver blood flow (*Panel A*) and protein binding expressed as free fraction of drug (*Panel B*). Three theoretical drugs with hepatic extraction ratios of 0.95, 0.3, and 0.1 are shown. Values are computed using the equation of Wilkinson & Shand (29).

Table 2 Effect of liver disease on the disposition of tranquilizers, sedatives, and analgesics

		$T_{\frac{1}{2}}$(hr)			C (ml/min)			Plasma protein binding (%)		
Drug	Ref.	Normal	Hepatitis	Cirrhosis	Normal	Hepatitis	Cirrhosis	Normal	Hepatitis	Cirrhosis
Tranquilizer										
Benzodiazepines										
Diazepam	25	32.7	74.5[a]	—	26.6	—	13.8[a]	97.8	—	95.3[a]
		46.6	—	105.6[a]	—	—	—	—	—	—
	33	31.1		164.0[a]	35.0		17.1[a]			
Chlordiazepoxide	37	26.3		62.3[a]	15.0		7.7[a]	96.4	—	94.6
Oxazepam	38	6.4	6.1	7.8	136	137	155	86.7	86.0	87.6
Lorazepam	39	22.1	23.8	31.9[a]	54.8	46.9	59.1	93.7	91.8[a]	88.6[a]
Meprobamate	40	12.6	21	25.5[a]	—	—	—	—	—	—
Chlorpromazine	41, 42	31	—	—	640	—	—	91–99	—	—
	43	31.0		15 no drugs		—	279	—	—	—
				26 on drugs	190	—	333	—	—	—
Barbiturates										
Phenobarbital	26	86	104	130[a]	—	—	—	—	—	—
Amobarbital	44	21.1		I = 39.4	34	—	I = 28	61	—41.5	
				II = 17.4			II = 42			
Hexobarbital	45	4.35	8.6[a]	—	3.57	1.94[a]	—	—	—	—
Pentobarbital	46	21.8	25.0	25.5	—	—	—	—	—	—
Narcotics/Analgesics										
Meperidine	27	3.37	6.99[a]	—	1261	649[a]	—	—	—	—
	28	3.20	—	7.0[a]	1316	—	664[a]	64	—	—
Morphine	14, 48	2.6	—	—	1800	—	—	35.1	—	25
Antipyrine	49, 50	12.0	19.5	33.8[a]	50.9	—	16.7[a]	< 10	—	—
	51	7.9		31.8[a]	58.6		18.5[a]			
Phenylbutazone	23	78	—	100+ [a]	—	—	—	—	—	—
	52	81.6		51	—	—	—	—	—	—

[a]Statistically significant. A + denotes multiple diagnoses including alcoholic or biliary cirrhosis, fibrosis, and chronic active hepatitis; group I had serum albumin values of < 3G%, group II > 3G%.

Diazepam (Valium) has been studied in relation to liver disease better than any other member of this class of drugs. It is first demethylated to desmethyldiazepam, which is then hydroxylated to oxazepam by the mixed-function oxygenase system (8, 9). These metabolites are pharmacologically active. In turn, oxazepam is conjugated and excreted in the urine as a glucuronide. In patients with alcoholic cirrhosis, acute viral hepatitis, and hepatic malignancy, but not in patients with extrahepatic obstruction, diazepam $T_{\frac{1}{2}}$ is prolonged (25, 33, 34). Along with the prolonged $T_{\frac{1}{2}}$, total plasma clearance of diazepam is decreased. In cirrhosis, protein binding is reduced which contributes to the increased volume of distribution and indicates that clearance of the unbound diazepam is further decreased compared to control values. Apart from the delayed elimination of the parent compound, the appearance in the plasma of the main metabolite desmethyldiazepam is slower in cirrhosis (25), and the elimination of the metabolite is impaired (35). This is consistent with earlier data obtained from rats with carbon tetrachloride–induced liver damage (25). The impairment of diazepam disposition is reversible, for in patients convalescing from viral hepatitis $T_{\frac{1}{2}}$ returns to normal, though later than the serum bilirubin or SGOT (25).

Since diazepam is commonly prescribed over a protracted period, Klotz et al (35) examined the effect of subchronic administration of diazepam on its disposition kinetics in human subjects without liver disease. They found that after six daily doses of 10 mg, diazepam $T_{\frac{1}{2}}$ increased from 33.9 hr to 52.9 hr. Moreover, there was an accumulation of desmethyldiazepam, which in later animal studies (35) was shown to inhibit the clearance of diazepam. In cirrhotic patients diazepam $T_{\frac{1}{2}}$ was prolonged (111.9 hr) after a single dose, consistent with previous findings, but after multiple dosing there was no further increase in $T_{\frac{1}{2}}$ (116.7 hr). It is possible that a reduced metabolite production resulted in concentrations inadequate to inhibit parent drug elimination (36). Further detailed pharmacokinetic studies of chronic diazepam administration in a larger number of liver disease patients and controls are needed to confirm these findings and determine whether this hypothesis is correct.

Like diazepam, chlordiazepoxide (Librium) has low intrinsic clearance and high protein binding (96%) in normal man. In cirrhotic patients the clearance and plasma protein binding are reduced, the volume of distribution is increased, and consequently the $T_{\frac{1}{2}}$ is strikingly prolonged (37). Moreover, the level of desmethylchlordiazepoxide in plasma is decreased in cirrhotic patients.

Oxazepam (Serax®) is conjugated by the liver to form a pharmacologically inactive glucuronide, which is excreted by the kidneys. In marked contrast to diazepam and chlordiazepoxide, the clearance, plasma protein binding, and urinary excretion of oxazepam in patients with cirrhosis or acute viral hepatitis remain normal (38). In addition, cirrhotic patients who had abnormal disposition of diazepam and meperidine had no difficulty metabolizing oxazepam.

Lorazepam (Ativan®), like oxazepam, is conjugated to form an inactive glucuronide, which is excreted by the kidneys. Studies in our laboratory show that the clearance of the parent compound is slightly, but not significantly, depressed in patients with cirrhosis and acute hepatitis, while plasma protein binding of the drug in these patients is significantly decreased. The $T_{\frac{1}{2}}$ of lorazepam (normally 22 hr) is unchanged in patients with hepatitis but is prolonged in those with cirrhosis, mainly due to increased V_d of the drug (39).

MEPROBAMATE In normal subjects the $T_{\frac{1}{2}}$ of meprobamate was 12.6 hr; in a patient with hepatitis, 21 hr; and in cirrhotic subjects, 25.5 hr. No detailed pharmacokinetic data are available (40).

CHLORPROMAZINE Chlorpromazine is a major tranquilizer of the phenothiazine group of drugs. Chlorpromazine $T_{\frac{1}{2}}$ values following an oral or intramuscular single dose or after multiple oral doses are similar: 30.5 hr, 31.3 hr, and 30.4 hr, respectively (4). Clearance has not been measured. Plasma protein binding in normal subjects varies from 91 to 99% (42). In cirrhotic subjects no prolongation of $T_{\frac{1}{2}}$ or decrease in C was demonstrated (43). The metabolites and plasma protein binding have not been measured.

Barbiturates

Phenobarbital has been proposed as a suitable sedative for patients with liver disease, because it is partially (30%) excreted by the kidneys in the unchanged form. The $T_{\frac{1}{2}}$ of phenobarbital, however, increased to 130 hr in cirrhotic patients compared to 86 hr in controls (26). This was accompanied by a 50% reduction in the urinary excretion of *p*-hydroxyphenobarbital. In subjects with viral hepatitis the disposition of phenobarbital was not significantly altered (26). The metabolism of other barbiturates (amobarbital (44), hexobarbital (45), and pentobarbital (46)) was also impaired in liver disease. The $T_{\frac{1}{2}}$ of hexobarbital was studied in patients with acute hepatitis during the acute phase and after apparent clinical and biochemical recovery, and it was noted that the $T_{\frac{1}{2}}$, although improved, remained abnormal (45).

Narcotics and Analgesics

Morphine, codeine, and methadone are opiates that are metabolized in the liver. The sedative effects of morphine are prolonged in animals with experimentally induced liver damage and in patients with liver disease (47). Detailed pharmacokinetic investigations of these drugs in the presence of liver disease are not available, but plasma protein binding of morphine has been shown to be reduced in liver disease (48). Meperidine (Demerol®) has been studied, and results show a significant prolongation of $T_{\frac{1}{2}}$ in acute viral hepatitis (27) and in cirrhosis (28), with a corresponding decrease in plasma clearance. Recovery from acute viral hepatitis results in the return of $T_{\frac{1}{2}}$ and C to normal values (27). Similarly, the elimination of antipyrine is impaired (49–51), but is reversible as the hepatitis resolves (51a). The studies of the effect of liver disease on phenylbutazone metabolism have been conflicting (23, 52), due in part to the concurrent use of other drugs in some patients (23).

CEREBRAL SENSITIVITY

In addition to abnormalities of drug disposition, there is increasing evidence that the cerebral sensitivity to many sedatives and analgesics is altered in patients with liver disease, particularly those with portasystemic encephalopathy. The fundamental mechanisms in the development of portasystemic encephalopathy are still poorly understood. However, it is well recognized that such patients are unduly sensitive to the effects of sedatives and narcotics. In early studies with morphine (53) and chlorpromazine (43, 54), a single, constant dose of either drug was given as a stress test, and the EEG was obtained at a standard time after drug administration. In controls and patients with liver disease, but not prior or current portasystemic encephalopathy, the EEG pattern was comparable. By contrast, patients with encephalopathy or prior encephalopathy, showed a significant alteration (increase in slow-wave activity in the 2–5 Hz region) following the drug administrion (53, 54). In similarly designed studies using amitriptyline (55) or diazepam (56), no adverse effects in such patients were noted. The interpretation of these negative studies, however, is difficult due to the arbitrary choice of dose and timing of the EEG. In a more recent study with diazepam in which the drug was infused slowly until the predetermined clinical end point of sedation was reached, there was evidence of increased cerebral sensitivity (57). The dosage requirement to induce the same clinical response in patients with liver disease was only two thirds of that required in normal subjects. Furthermore, at the time of peak response, which was at the end of the drug infusion, patients with prior encephalopathy showed a significant alteration in the EEG, and this correlated with the presence of an abnormal baseline EEG and the degree to which the serum albumin was lowered. The correlation between the EEG changes and the low serum albumin implies either that patients with more severe liver disease (reflected by the low serum albumin) had greater cerebral sensitivity, or that the lower serum albumin permitted more active drug to reach the cerebral receptor sites. Within one hour, the EEG of all subjects had returned to baseline. A further example of the importance of experimental design is the observation that a single dose of the monoamine oxidase inhibitor, tranylcypromine, had no effect on the clinical status or EEG of patients with prior portasystemic encephalopathy, while chronic administration of the drug had marked effect (55).

There is essentially no information available concerning cerebral sensitivity of patients with hepatocellular liver disease to the other available sedatives and analgesics. Studies of this nature, which would utilize the newer EEG techniques that employ a power spectrum analysis of a conventional EEG recording, are now feasible, and should provide a more quantitative and sensitive assessment. In the interim it is evident that many sedatives and analgesics may cause a deterioration of cerebral function and alteration of EEG pattern in patients with liver disease who are prone to portasystemic encephalopathy; this finding should raise a strong note of caution in the therapy of such patients, regardless of the drug elimination pattern.

THE IDEAL SEDATIVE OR ANALGESIC

In addition to altered cerebral sensitivity many factors are important in determining the safety and risks of sedatives or analgesics in patients with liver disease. These include drug disposition (already discussed at length), plasma binding, which affects drug elimination and determines the concentration of the unbound drug fraction at receptor sites, and finally the therapeutic/toxic ratio of the drug. In view of these considerations the ideal agent for sedation or analgesia in patients with liver disease should have a rapid titratable effect, short elimination half-life, and a wide margin of safety with minimal or no risk of depressing vital functions. Its elimination or plasma binding should not be affected by liver disease. The drug should have no enhanced effect on the site of action, and there should be no accumulation of active metabolites. Table 3 attempts to assess a number of drugs against these criteria. It is apparent that no single agent is ideal for use as a sedative or analgesic, although oxazepam meets most of the criteria, and on pharmacokinetic grounds it would seem to be a reasonable tranquilizer to use in patients with liver disease. However, it has the disadvantage of not being available for parenteral administration, and its effects in terms of cerebral sensitivity require study. While there is no ideal tranquilizer/sedative or analgesic, it is possible to make a rational choice of drug based on the specific clinical needs to be met and on available pharmacokinetic information.

GENERAL GUIDELINES

In the treatment of patients with renal failure it is possible to arrive at a proper drug dose based on the assessment of creatinine clearance. To achieve a similar degree of sophistication in the treatment of patients with liver disease, a reliable predictive test to assess the degree of hepatic drug metabolism must first be

Table 3 Assessment of tranquilizers, sedatives, and analgesics for use in liver disease[a]

Criteria	Diazepam	Chlordiazepoxide	Oxazepam	Lorazepam	Phenobarbital	Meperidine
1. Rapid titratable effect	√	√	X			√
2. No depression of vital function	√	√			X	X
3. Short $T_{\frac{1}{2}}$	X	X	√	X	X	√
4. Elimination not affected by liver disease	X	X	√	√	X	X
5. Protein binding not affected by liver disease	X	X	√	X		√
6. No enhanced effect on receptor	X					
7. No accumulation of active metabolite	X	X	√	√	√	X

[a] X indicates a false or unfavorable statement; √ indicates a true or favorable statement; and blank spaces reflect lack of adequate data or uncertainty about result.

devised. At present no such test is available. The conventional liver "function" tests bear no consistent relation to the liver's capacity to metabolize drugs. For amobarbital (44), antipyrine (49, 58), and phenylbutazone (23), low serum albumin correlates with impaired elimination, but abnormal elimination can also occur despite a normal albumin concentration. For some other sedatives and analgesics (i.e. meperidine (27, 28) and chlordiazepoxide (37)) no correlation between these liver tests and drug elimination was detected. Some investigators have suggested that the elimination of model drugs such as galactose (59) or indocyanine green (34) or antipyrine (58) may serve as a guide or prototype for the elimination of similar agents. Unfortunately the correlation between different drug eliminations shows considerable variability, which makes the application of such information to the individual patient difficult. Thus, the need for a more precise predictive test(s) is obvious, but its availability must await future developments. In the meantime physicians must rely solely on clinical guidelines, or in difficult cases on monitoring plasma drug levels when available. The advantages and limitations of each agent should be recognized and the indications carefully scrutinized. When use of a drug whose metabolism may be impaired by liver disease is essential or unavoidable, the dose should be adjusted based on pharmacokinetic information, and it should be carefully titrated according to the clinical response.

CONCLUDING REMARKS

The elimination of many sedatives and analgesics is impaired in patients with hepatocellular disease, and preliminary studies suggest that such individuals have increased cerebral sensitivity to these agents. It is not surprising therefore that these drugs constitute a common precipitant of coma in individuals with severe liver disease (1). It is also conceivable that sedatives given to patients with significant hepatic dysfunction may impair their psychomotor performance in subtle ways and even increase the risk of accident at work or on the road. It is evident from this brief review that much remains to be elucidated. High priority areas for future investigation relate to: (*a*) further clarification of the disposition kinetics of other sedatives and analgesics, to search for drugs handled normally in the presence of liver disease; (*b*) assessment of cerebral sensitivity to sedatives by sophisticated EEG techniques; (*c*) determination of sedative metabolism in patients given the drugs over a prolonged period; (*d*) attempts to elaborate a test predicting hepatic drug elimination; and (*e*) correlation of drug elimination in vivo with its hepatic metabolism in vitro. Answers to these questions are needed to improve our understanding and management of patients with liver disease.

ACKNOWLEDGMENTS

We wish to thank Dr. Grant R. Wilkinson for his helpful suggestions in the preparation of this article.

Literature Cited

1. Fessel, J. M., Conn, H. O. 1972. An analysis of the causes and prevention of hepatic coma. *Gastroenterology* 62:191
2. Schenker, S., Hoyumpa, A. M., Wilkinson, G. R. 1975. *Med. Clin. North Am.* 59:887–96
3. Fingl, E., Woodbury, D. M. 1975. General Principles. In *The Pharmacological Basis of Therapeutics*, ed. L. S. Goodman, A. Gilman, pp. 1–46. New York: MacMillan
4. Estabrook, D. W., Matsubara, T., Mason, J. I., Werringloer, J., Baron, J. 1973. Studies on the molecular function of cytochrome P450 during drug metabolism. *Drug Metab. Dispos.* pp. 98–110
5. Schwartz, M. A., Postma, E. 1966. Metabolic N-Demethylation of chlordiazepoxide. *J. Pharm. Sci.* 55:1358–62
6. Schwartz, M. A., Postma, E., Gaut, Z. 1971. Biological half-life of chlordiazepoxide and its metabolite, Demoxepam, in man. *J. Pharm. Sci.* 60:1500–3
7. De Silva, J. A. F., Schwartz, M. A., Stefanovic, V., Kaplan, J., D'Arconte, L. 1964. Determination of Diazepam (Valium) in blood by gas liquid chromatography. *Anal. Chem.* 36:2099
8. Schwartz, M. A., Koechlin, B. A., Postma, E., Palmer, S., Krol, G. 1965. Metabolism of diazepam in rat, dog and man. *J. Pharmacol. Exp. Ther.* 149:423–35
9. Randall, L. O., Scheckel, C. L., Banziger, R. F. 1965. Pharmacology of the metabolites of chlordiazepoxide and diazepam. *Curr. Ther. Res. Clin. Exp.* 7:590–606
10. Sulser, F., Watts, J., Brodie, B. B. 1962. On the mechanism of antidepressant action of impraminelike drugs. *Ann. NY Acad. Sci.* 96: 279–88
11. Harvey, S. C. 1975. Hypnotics and sedatives. See Ref. 3, pp. 102–23
12. Vessman, J., Alexanderson, B., Sjogvist, B., Strindberg, B., Sundwall, A. 1973. Comparative pharmacokinetics of oxazepam and nortriptyline after simple oral doses in man. In *The Benzodiazepines*, ed. S. Garattini, E. Mussini, L. O. Randall, p. 165. New York: Raven
13. Elliott, H. W. 1976. Metabolism of lorazepam. *Br. J. Anaesth.* 48: 1017–23
14. Brunk, S. F., Delle, M. 1974. Morphine metabolism in man. *Clin. Pharmacol. Ther.* 16:51–57
15. Vessel, E. S. 1972. Pharmacogenetics. *N. Engl. J. Med.* 287:904–9
16. Crooks, J., O'Malley, K., Stevenson, I. H. 1976. Pharmacokinetics in the elderly. *Clin. Pharmacokinetics* 1:280–96
17. O'Malley, K., Crooks, J., Duke, H., Stevenson, I. H. 1971. Effect of age and sex on human drug-metabolism. *Br. Med. J.* 3:607–9
18. Baselt, R. C., Casarett, L. J. 1972. Urinary excretion of methadone in man. *Clin. Pharmacol. Ther.* 13: 64–70
19. Basu, T. K., Dickerson, J. W. T. 1974. Interrelationships of nutrition and the metabolism of drugs. *Chem. Biol. Interactions* 8:193–206
20. Alvares, A. P., Anderson, K. E., Conney, A. H., Kappas, A. 1976. Interactions between nutritional factors and drug biotransformations in man. *Proc. Natl. Acad. Sci. USA* 73:2501–4
21. Kappas, A., Anderson, K. E., Conney, A. H., Alvares, A. P. 1976. Influence of dietary protein and carbohydrate on antipyrine and theophylline metabolism in man. *Clin. Pharmacol. Ther.* 20:643–53
22. Conney, A. H., Pantuck, E. J., Hsiao, K. C., Garland, W. A., Anderson, K. E., Alvares, A. P., Kappas, A. 1976. Enhanced phenacetin metabolism in human subjects fed charcoal-broiled beef. *Clin. Pharmacol. Ther.* 20:633–42
23. Levi, A. J., Sherlock, S., Walker, P. 1968. Phenylbutazone and isoniazid metabolism in patients with liver disease in relation to previous drug therapy. *Lancet* 1:1275–79
24. Levy, G. 1977. Pharmacokinetics in renal disease. *Am. J. Med.* 62: 461–65
25. Klotz, U., Avant, G. R., Hoyumpa, A., Schenker, S., Wilkinson, G. R. 1975. The effect of age and liver disease in the disposition and elimination of diazepam in adult man. *J. Clin. Invest.* 55:347–59
26. Alvin, J., McHorse, T., Hoyumpa, A., Bush, M. T., Schenker, S. 1975. The effect of liver disease in man on the disposition of pheno-

barbital. *J. Pharmacol. Exp. Ther.* 192:224–35

27. McHorse, T. S., Wilkinson, G. R., Johnson, R. F., Schenker, S. 1975. Effect of acute viral hepatitis in man on the disposition and elimination of meperidine. *Gastroenterology* 68: 775–80
28. Klotz, U., McHorse, T. S., Wilkinson, G. R., Schenker, S. 1974. The effect of cirrhosis on the disposition and elimination of meperidine. *Clin. Pharmacol. Ther.* 16:667–675
29. Wilkinson, G. R., Shand, D. G. 1975. A physiological approach to hepatic drug clearance. *Clin. Pharmacol. Ther.* 18:377–90
30. Branch, R. A., Shand, D. G. 1976. Propranolol disposition in chronic liver disease: A physiological approach. *Clin. Pharmacokinetics* 1:264–79
31. Wilkinson, G. R. 1976. Pharmacokinetics in disease states modifying body perfusion. In *The Effect of Disease States on Drug Pharmacokinetics*, ed. L. Z. Benet, pp. 13–32. Washington, DC: Pharm. Assoc./Acad. Pharm. Sci.
32. Branch, R. A., Shand, D. G. 1976. Hepatic drug clearance in chronic liver disease. See Ref. 31, pp. 77–86
33. Andreasen, P. B., Hendel, J., Greisen, G., Hvidberg, E. F. 1976. Pharmacokinetics of diazepam in disordered liver function. *Eur. J. Clin. Pharmacol.* 10:115–20
34. Hepner, G. W., Vessell, E. S., Lipton, A., Harvey, H. A., Wilkinson, G. R., Schenker, S. 1977. Disposition of aminopyrine, antipyrine, diazepam and indocyanine green in patients with liver disease or on anticonvulsant drug therapy. *J. Lab. Clin. Med.* 90:440–56
35. Klotz, U., Antonin, K. H., Bieck, P. R. 1976. Comparison of the pharmacokinetics of diazepam after single and subchronic doses. *Eur. J. Clin. Pharmacol.* 10:121–26
36. Klotz, U., Antonin, K. H., Brugel, H., Bieck, P. R. 1977. Disposition of diazepam and its major metabolite desmethyldiazepam in patients with liver disease. *Clin. Pharmacol. Ther.* 21: 430–36
37. Roberts, R., Wilkinson, G. R., Branch, R., Schenker, S. 1977. Effect of age and cirrhosis on the disposition and elimination of chlordiaze poxide. *Gastroenterology.* 73:A45/1243
38. Shull, H. J. Jr., Wilkinson, G. R., Johnson, R., Schenker, S. 1976. Normal disposition of oxazepam in acute viral hepatitis and cirrhosis. *Ann. Intern. Med.* 84:420–25
39. Kraus, J. W., Marshall, J. P., Johnson, R., Wilkinson, G. R., Schenker, S. 1977. Lorazepam elimination in liver disease. *Gastroenterology.* 73:A30/1228
40. Held, H., Von Oldershausen, H. F. 1969. Zur pharmakokinetic von meprobamat bei chronischen hepatopathien und arzneimittelsucht. *Klin. Wochenschr.* 47:78–80
41. Dahl, S. G., Strandjord, R. E. 1977. Pharmacokinetics of chlorpromazine after single and chronic dosage. *Clin. Pharmacol. Ther.* 21:437–48
42. Curry, S. H. 1970. Plasma protein binding of chlorpromazine. *J. Pharm. Pharmacol.* 22:193–97
43. Maxwell, J. D., Carrella, M., Parkes, J. D., Williams, R., Mould, G. P., Curry, S. H. 1972. Plasma disappearance and cerebral effects of chlorpromazine in cirrhosis. *Clin. Sci.* 43:143–51
44. Mawer, G. E., Miller, N. E., Turnberg, L. A. 1972. Metabolism of amylobarbitone in patients with chronic liver disease. *Br. J. Pharmacol.* 44:549–60
45. Breimer, D. D., Zilly, W., Richter, E. 1975. Pharmacokinetics of hexobarbital in acute hepatitis and after apparent recovery. *Clin. Pharmacol. Ther.* 18:433–40
46. Held, H., Von Oldershausen, H. F., Remmer, H. 1970. Der abbau von pentobarbital bei leberschaden. *Klin. Wochenschr.* 48:565–67
47. Hoyumpa, A., 1976. Diagnosis and management of hepatic encephalopahy. In *Hepatic Support in Acute Liver Failure*, ed. G. G. R. Kuster, pp. 88–124. Springfield, Ill: Thomas
48. Olsen, G. D., Bennett, W. M., Porter, G. A. 1975. Morphine and phenytoin binding to plasma proteins in renal and hepatic failure. *Clin. Pharmacol. Ther.* 17:677–84
49. Branch, R. A., Herbert, C. M., Read, A. E. 1973. Determinants of serum antipyrine half-life in patients with liver disease. *Gut* 14:569–73
50. Andreasen, P. B., Ranek, L. 1975. Liver failure and drug metabolism.

Scand. J. Gastroenterol. 10:293–97

51. Andreasen, P. B., Ranek, L., Statland, B. E., Tygstrup, N. 1974. Clearance of antipyrine-dependence of quantitative liver function. *Eur. J. Clin. Invest.* 4:129–34

51a. Burnett, D. A., Barak, A. J., Tuma, D. J., Sorrell, M. F. 1976. Altered elimination of antipyrine in patients with acute viral hepatitis. *Gut* 17:341–44

52. Hvidberg, E. E., Andreasen, P. B., Ranek, L. 1974. Plasma half-life of phenylbutazone in patients with impaired liver function. *Clin. Pharmacol. Ther.* 15:171–77

53. Laidlaw, J., Read, A. E., Sherlock, S. 1961. Morphine tolerance in hepatic cirrhosis. *Gastroenterology* 40:389–96

54. Read, A. E., Laidlaw, J., McCarthy, C. F. 1969. Effects of chlorpromazine in patients with hepatic disease. *Br. Med. J.* 3:497–99

55. Morgan, M. H., Read, A. E. 1972. Antidepressants and liver disease. *Gut* 13:697–701

56. Murray-Lyon, I. M., Young, J., Parkes, J. D., Knill-Jones, R. P., Williams, R. 1971. Clinical and electroencephalographic assessment of diazepam in liver disease. *Br. Med. J.* 4:265–66

57. Branch, R. A., Morgan, M. H., James, J., Read, A. E. 1976. Intravenous administration of diazepam in patients with chronic liver disease. *Gut* 17:975–83

58. Branch, R. A., James, J. A., Read, A. E. 1976. The clearance of antipyrine and indocyanine green in patients with chronic liver disease. *Clin. Pharmacol. Ther.* 20:81–89

59. Ranek, L., Andreasen, P. B., Tygstrup, N. 1976. Galactase elimination capacity as a prognostic index in patients with fulminant liver failure. *Gut* 17:959–64

Ann. Rev. Med. 1978. 29:219–29
Copyright © 1978 by Annual Reviews Inc. All rights reserved

SPECIFICITY OF DEPRESSIVE DISEASES

◆7278

Donald R. Sweeney, M.D., Ph.D. and James W. Maas, M.D.
Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06510

INTRODUCTION

The syndrome of depression has been recognized for centuries. However, only recently has there been a coalescence of nosological refinement and biotechnical advances sufficient to generate promising leads toward identifying the substrates of the depressive disorders.

Because depressive states are characterized by prominent dysphoria, often with accompanying feelings of guilt, anxiety, or hopelessness, depression (along with mania) is commonly categorized as an affective disorder. However, depression may also include cognitive, motivational, and neurovegetative changes (1). Cognitive changes include memory and concentration difficulties and, in some patients, psychotic ideation. Motivational disturbances are described by terms such as lassitude or anergy. Neurovegetative changes include alterations in sleep patterns, appetite, and sexual activities as well as psychomotor signs such as agitation or retardation. These changes may lead to social withdrawal, despair, and preoccupation with suicide.

It is generally thought that depression is a psychobiological disorder, i.e. both psychological and biological factors underlie the pathogenesis of most depressive states, and many of the formulations of the genesis of this disorder have been focused upon psychological dimensions (2–7). In recent years increasing attention has been paid to genetic (8) and nosological questions as well as to investigations of the metabolism and/or disposition of neurotransmitters, particularly norepinephrine and serotonin (9, 10). The information from these studies suggests that the diagnosis of depression includes a heterogeneous group of disorders, and that it may be possible to identify subtypes through use of specific criteria.

The present discussion is confined to a selective review of recent contributions to the nosology of depression and to studies of neurotransmitter metabolism in depressed patients. Emphasis is placed on clinical studies with the hope that the literature cited will guide those who wish to review in detail the

0066-4219/78/0401-0219$01.00

laboratory research in these areas. Space limitations prevent the detailed exploration of the important work represented by twin and family history studies, enzyme studies, and neurophysiological studies.

DIAGNOSTIC CONSIDERATIONS

Some of the earlier schemes for diagnosing and differentiating the affective disorders have proved to be less useful or less reliable than others. For example, the current *Diagnostic and Statistical Manual of Mental Disorders* (DSM-II) (11), soon to yield to DSM-III, incorporates the distinction between "endogenous" and "reactive" depression in differentiating "psychotic depressive reaction" from "major affective disorder." Although there is some evidence to support an endogenous-reactive subtyping (12), the presence or absence of an environmental precipitant has not been a criterion by which homogeneous subgroups of depressed patients can be identified. Environmental stress may trigger relapses in patients whose previous depressive episodes have been endogenous, while reactive depressed patients may progress to episodes that appear unrelated to any precipitants. It has been suggested that the terms endogenous and reactive are useful as descriptors, but not as major, independent categories of depression (13). DSM-II also distinguishes between neurotic and psychotic depression, based on the extent to which the depression is accompanied by loss of contact with reality. Psychotic depressions are generally more severe, being accompanied frequently by the full range of symptoms and signs of depression. However, some neurotically depressed patients may have severe, debilitating illness without psychotic symptoms, while others may have a chronic depression of low intensity, which is sometimes labeled characterological. As with the endogenous-reactive dimension, there is no evidence that the psychotic-neurotic dimension forms the basis for identification of homogeneous depressive subtypes.

Because of the general lack of utility of these older diagnostic groupings of depression, efforts have been made to develop more reliable classification schemes that will separate subtypes of depression with genetic or biochemical similarities. Two such diagnostic schemes are the primary-secondary affective-disorder concept and the bipolar-unipolar dichotomy.

The differentiation between primary and secondary affective disorder was developed by the Washington University group (14, 15) and subsequently incorporated into the Research Diagnostic Criteria (RDC) (16). The RDC are reliable operational diagnostic criteria designed to allow investigators to form clinically homogeneous groups of patients (17). A primary affective disorder[1] is a depressive episode characterized by marked dysphoria along with cognitive, motivational, and neurovegetative changes in a patient with no evidence of illness other than depression. Included in the category of secondary affective disorder are those patients with depressive symptoms preceded by or associated with other illnesses (e.g. other psychiatric disorders such as schizophrenia or substance abuse, or medical illnesses such as chronic cardiopulmonary diseases or diabetes).

---

[1] A depressive illness categorized as primary affective disorder by the Washington University criteria is relabeled as "major primary depressive disorder" in the RDC.

This differentiation thereby places affective disorders with possibly more heterogeneous origins in a separate category (secondary affective disorder). There is no a priori reason to believe that patients in the primary affective category represent a homogeneous group, but this categorization provides a relatively good starting point for the investigator of the psychobiological correlates of depression (see next section).

The differentiation of unipolar and bipolar affective disorders has also proved to be important. A depressed patient with a history of manic or hypomanic episodes is considered to be suffering from a bipolar depression, while the designation of unipolar depression is given to those patients who do not have a history of mania. The symptoms and signs of mania include elevated or irritable mood, increased physical and social activity, pressured speech, grandiosity, decreased sleep, and poor social judgment. Unipolar and bipolar depression have been found to differ in clinical (18, 19) and neurophysiological dimensions (20, 21). Family history studies have reported that bipolar patients have more bipolar relatives than do unipolar patients (e.g. 22–24). Lithium carbonate has been reported to exert an antidepressant effect in bipolar depressed patients but not in unipolar depressed patients (25–27).

It would thus appear that the bipolar-unipolar categories separate patients into groups that have some psychobiological homogeneity. It should be emphasized, however, that in none of the studies cited above is there complete independence of unipolar and bipolar subtypes, i.e. in most studies, regardless of the variable being considered, there is overlap between the unipolar and bipolar groups. Nevertheless, clinical studies of affective disorder can no longer be considered complete without careful specification of polarity and a separate consideration of data collected from the two groups.

NEUROTRANSMITTERS

Certain of the central nervous system (CNS) neurotransmitter substances, sometimes called biogenic amines, have been important in the attempts to identify the biological substrates of depression. The neurotransmitters that have occupied much of the attention of clinical investigators are norepinephrine, a catecholamine, and serotonin, an indoleamine. Interest in these amines was stimulated several years ago when it was observed that drugs that deplete them centrally (e.g. reserpine, used to treat hypertension) produced depression in some patients. Conversely, drugs that increase the amounts of these amines in the CNS (e.g. iproniazid, used to treat tuberculosis) produced euphoria or psychological activation in some patients. Subsequently, investigators have been interested in evaluating the extent to which these central amines are involved in depressive illness. In clinical studies one approach has been to study amine metabolites in the body fluids of depressed patients. This approach is indirect, and there are some major unresolved issues regarding the extent to which these metabolites reflect central events. However, some intriguing findings have emerged from this approach and these are discussed below.

Norepinephrine

Noradrenergic neurons have their cell bodies in the brain stem, with an especially high density in a pontine nucleus called the locus ceruleus. Noradrenergic axons ascend in the medial forebrain bundle and give off terminals to subcortical areas, such as the hypothalamus, and to the entire cerebral cortex. There are also other noradrenergic pathways to the cerebellum and spinal cord. The major CNS metabolite of norepinephrine is 3-methoxy-4-hydroxyphenethyleneglycol (MHPG) (28–31). Animal studies have shown that both brain norepinephrine and MHPG increase with stress or electrical stimulation and decrease with destruction of the nucleus locus ceruleus (32, 33). There is also some evidence that a significant fraction of the MHPG found in urine originates in brain (34–36); hence, measurement of urinary MHPG may provide an estimate of central noradrenergic metabolism and/or disposition in depressed patients. This possibility has led to a series of studies of urinary MHPG in depressed and normal subjects, and these are briefly reviewed below.

The typical, well-controlled study of urinary MHPG excretion involves three or more complete 24 hour urine collections, usually taken at least several days after a patient has been admitted to the hospital. Patients who have shown significant improvement in depression during the early period of hospitalization are usually excluded from these studies, thereby eliminating those whose symptoms are not enduring. For at least several days prior to urine collection patients are maintained on a diet that excludes foods known to contribute to the urinary pool of catecholamine metabolites.

In a study of diagnostically heterogenous depressed patients and healthy comparison subjects, it was found that the patients excreted significantly lower amounts of urinary MHPG than the healthy comparison subjects (37). There was, however, considerable overlap between the two groups. In subsequent studies incorporating greater diagnostic specificity, it was found that those depressed patients diagnosed as having a primary affective disorder (see earlier section) were particularly likely to excrete less MHPG than comparison subjects (38, 39). The differences between the two groups were not absolute, however.

Another set of studies has focused on MHPG excretion by bipolar depressed patients. It has been found that this group excretes less MHPG than characterologically (chronic, environmentally reactive, and usually non-drug responsive) depressed patients (40) or healthy comparison subjects (41–43).

Urinary MHPG has also been shown, in three separate studies, to predict differential treatment response to tricyclic antidepressants. Patients with low urinary MHPG respond favorably to imipramine, but not to amitriptyline, while patients with normal or high urinary MHPG respond favorably to amitriptyline, but not to imipramine (44–46). Moreover, patients with low urinary MHPG experience an elevation of mood, while those with normal or high MHPG experience no change in mood, when given brief trials of dextroamphetamine (47, 48).

These pharmacological and biochemical differences between depressed patients may represent evidence that depression is a biochemically heterogeneous disorder. It has been suggested that two depressive subgroups may be distinguished on the basis of these biochemical and pharmacological criteria (9). One subgroup (group A) excretes low urinary MHPG, has a brightening of mood in response to dextroamphetamine, and a positive treatment response to imipramine but not amitriptyline. The other subgroup (group B) excretes normal or high amounts of MHPG, exhibits negligible dextroamphetamine-induced mood change, and responds positively to amitriptyline but not imipramine.

Because a decrement in urinary MHPG may reflect a functional change in central norepinephrine disposition, and because both imipramine and dextroamphetamine increase the functional availability of norepinephrine by blocking its reuptake into the presynaptic cell, it has been postulated that patients in group A have a disorder involving norepinephrine metabolism and/or disposition. Since patients in group B have normal or high urinary MHPG, and because amitriptyline produces a reuptake blockade of serotonin, but not of norepinephrine, these patients do not appear to have a norepinephrine-related disorder, but it is possible that alterations in the disposition of serotonin are involved in their illness (9).

In summary, there is some initial evidence for biochemical and pharmacological separability of depressed patients on the basis of indirect indices of central noradrenergic function. However, it cannot be assumed that a decrement in urinary MHPG levels is specific to depressive disorders. In a recent study of urinary MHPG in schizophrenia, it was found that, while these patients as a group were not different from comparison subjects, some schizophrenic patients had MHPG values lower than any of the depressed patients (39). Phrased differently, a low urinary MHPG value is predicted by a diagnosis of primary affective disorder, and this prediction is even stronger for a primary affective patient of the bipolar subtype. However, if we are given a patient with low urinary MHPG, we cannot reliably predict the psychiatric diagnosis.

Serotonin

Serotonergic neurons have their cell bodies in the raphe nuclei, which are located in the midbrain and upper pons. Axons ascend in the medial forebrain bundle giving off terminals diffusely in many parts of the brain, including the cerebral cortex, with especially dense terminals in hypothalamic areas. Clinical studies of serotonin metabolism in depressed patients have focused on the measurement of cerebrospinal fluid (CSF) concentration of 5-hydroxyindoleacetic acid (5-HIAA), the principal metabolite of serotonin. There are two major approaches to measuring CSF 5-HIAA. The first involves simply collecting the CSF sample and performing the assay using spectrofluorometric methods or, more recently, mass fragmentography. The other approach involves the administration of oral probenecid, usually beginning about 24 hours before the lumbar puncture. Probenecid interferes with the efflux of organic acid metabolites such as 5-HIAA from the CNS. Probenecid administration is thought to facilitate the estimation of the turnover of brain amines such as serotonin. There is some controversy regarding the extent to which lumbar concentrations of 5-HIAA reflect brain serotonin metabolism, but it is probable that some portion of lumbar 5-HIAA originates in brain (49, 50).

There have been several clinical studies in which CSF 5-HIAA has been compared in depressed patients and comparison subjects. Because of the difficulties in obtaining healthy individuals as volunteers for lumbar puncture studies, comparison groups in these studies have usually consisted of neurological patients. 5-HIAA has been reported to be lower in a diagnostically heterogeneous group of depressed patients than in comparison subjects (51), lower in primary depressives than in comparison subjects (52, 53), and lower in endogenous depressives after probenecid than in comparison subjects (54). However, other studies have failed to confirm these results both with and without the administration of probenecid (55, 56).

Of substantial interest to the present discussion is a recent study employing mass fragmentographic assay techniques (57). This study reported that values for CSF 5-HIAA obtained from 68 patients with "severe primary depressive illness" were distributed in a lower mode with 29 per cent of the patients and an upper mode with the remainder of the patients, i.e. the values for CSF 5-HIAA appeared to be distributed bimodally. This distribution was not related to demographic or pharmacologic variables. A similar bimodal distribution was suggested by an earlier study of probenecid-induced accumulation of 5-HIAA (54). These findings suggest that there may be a distinct biochemical subgroup of depressed patients characterized by low CSF 5-HIAA. As with studies on MHPG then, there is evidence that some, but not all, depressed patients may have some alteration of serotonin metabolism and/or disposition as reflected by CSF 5-HIAA concentrations. The MHPG data are also paralleled by the fact that nondepressed psychiatric patients have also been reported to have low CSF 5-HIAA values (55, 56). While there is evidence, as noted above, suggesting that there are identifiable subtypes of depression that may be related to alterations in noradrenergic or serotonergic systems, it may be that such single-amine theories are too simple, and it is important to examine these systems as they may interact with each other.

Norepinephrine-Serotonin Interactions

The notion of norepinephrine-serotonin interactions in depression is not a new one. There is some evidence that CSF 5-HIAA is not state dependent, e.g. reduced CSF 5-HIAA has been reported to occur in both manic and depressed patients (52, 53). Moreover, CSF 5-HIAA has been observed, in some studies, to remain low after remission of depressive symptoms (51–53). In contrast, the weight of evidence supports the possibility that urinary MHPG is state dependent. In rapidly cycling bipolar patients, urinary MHPG is significantly higher during mania than during depression (58–61). Urinary MHPG also varies, within relatively narrow limits, with state anxiety (62, 63). Furthermore, data from a recent longitudinal study indicate that patients who have recovered from serious depressions show significant increases in urinary MHPG, while patients with recurrence of depression maintain their initial low levels of MHPG (D. Pickar, D. R. Sweeney, J. W. Maas, and G. R. Heninger, submitted for publication).

The notion that lower CSF 5-HIAA may represent a trait and urinary MHPG a state has led to a permissive hypothesis, which relates biogenic amines to depression (64, 65). This hypothesis states that some patients have a deficiency in serotonin metabolism and/or disposition, which permits the emergence of depression or mania when changes in noradrenergic systems occur.

In contrast to the permissive hypothesis, which postulates an interaction between serotonin and norepinephrine, is the possibility reviewed above that some patients have norepinephrine depressions and other patients have serotonin depressions. This hypothesis would predict that depressed patients would have either low CSF 5-HIAA or low urinary MHPG but not both. There is one recent study that supports this idea: 80 depressed patients diagnosed as having a primary affective disorder, and subcategorized as unipolar or bipolar, were found to have a bimodal distribution of probenecid-induced accumulation of CSF 5-HIAA. Urinary MHPG in the low 5-HIAA group was significantly higher than in the high 5-HIAA group. Unipolar and bipolar patients were not distributed differently in the two 5-HIAA modes (66).

At present, data are insufficient to support either an interactive or an independent hypothesis of alterations in central norepinephrine and serotonin systems in depression. Further studies are needed in which, after careful diagnosis, depressed patients are studied longitudinally with measurement of both norepinephrine and serotonin metabolites during drug-free and treatment periods.

SUMMARY AND CONCLUSIONS

The issue of specificity in the affective disorders (mania and depression) has been reviewed, and the available data suggest the following:

1. The use of a new diagnostic category, i.e. primary affective disorder, allows clinical investigators to reliably identify patients having relatively homogeneous signs and symptoms. While this grouping does not imply that patients so diagnosed are biologically homogeneous, it is useful as a starting point in biological studies of the affective disorders.
2. The further division of primary affective disorder into unipolar and bipolar subtypes has resulted in data that suggest that there are clinical, biochemical, and genetic differences between these subgroups.
3. There is evidence indicating that urinary MHPG and CSF 5-HIAA may reflect the metabolism and/or disposition or CNS norepinephrine and sero-

tonin, respectively. For this reason clinical investigators have focused on urinary MHPG and CSF 5-HIAA in studies aimed at elucidating which CNS amines may be related to mania or depression.

4. There is a subgroup of depressed patients who have low urinary MHPG values and respond favorably to treatment with imipramine but not amitriptyline. It is suggested that this subgroup may have an alteration in the functioning of brain noradrenergic systems.
5. There is evidence that indicates that CSF 5-HIAA values in depressed patients are bimodally distributed, and this has also been taken as evidence for the biochemical heterogeneity of the depressive disorders.
6. Overall, the available evidence supports the likelihood that affective disorders are a heterogeneous group of illness having diverse types of pathophysiology.

Literature Cited

1. Beck, A. T. 1967. *Depression: Clinical, Experimental and Theoretical Aspects*, pp. 10–39. New York: Harper & Row. 370 pp.
2. Freud, S. 1948. Mourning and melancholia. In *Collected Papers*, Vol. 4, pp. 152–72. London: Hogarth. 508 pp.
3. Bibring, E. 1965. The mechanism of depression. In *Affective Disorders*, ed. P. Greenacre, pp. 13–48. New York: Int. Univ. Press. 212 pp.
4. Lazarus, A. A. 1968. Learning theory and the treatment of depression. *Behav. Res. Therap.* 6:83–39
5. Costello, C. G. 1972. Depression: Loss of reinforcement or loss of reinforcer effectiveness? *Behav. Therap.* 3:240–47
6. Seligman, M. E. P. 1972. Learned helplessness. *Ann. Rev. Med.* 23:407–12
7. Eastman, C. 1976. Behavioral formulations of depression. *Psychol. Rev.* 83:277–91
8. Gershon, E. S., Bunney, W. F. Jr., Leckman, J. F., Van Eerdewegh, M., DeBauche, B. A. 1976. The inheritance of affective disorders: A review of data and of hypotheses. *Behav. Genet.* 6:227–61
9. Maas, J. W. 1975. Biogenic amines and depression. *Arch. Gen. Psychiatry* 32:1357–64
10. Coppen, A. 1972. Indoleamines and affective disorders. *J. Psychiatr. Res.* 9:163–71
11. Diagnostic and Statistical Manual of Mental Disorders. 1968. Washington, DC: Am. Psychiatr. Assoc. 134 pp.
12. Mendels, J., Cochrane, C. 1968. The nosology of depression: The endogenous-reactive concept. *Am. J. Psychiatry* (Suppl.) 124:1–10
13. Klein, D. F. 1974. Endogenomorphic depression: A conceptual and terminological revision. *Arch. Gen. Psychiatry* 31:447–54
14. Winokur, G. 1973. The types of affective disorders. *J. Nerv. Ment. Dis.* 156:82–96
15. Feighner, J. P., Robins, E., Guze, S. B., Woodruff, B. A., Winokur, G., Munoz, R. 1972. Diagnostic criteria for use in psychiatric research. *Arch. Gen. Psychiatry* 26:57–63
16. Spitzer, R., Endicott, J., Robins, E. 1975. *Research Diagnostic Criteria.* New York: NY State Dep. Ment. Hygiene, NY State Psychiatr. Inst., Biom. Res. 31 pp.
17. Spitzer, R., Endicott, J., Robins, E. 1977. Using the research diagnostic criteria to study affective disorders. *Sci. Proc. Am. Psychiatr. Assoc.* 93:55 (Abstr.)
18. Perris, C. 1969. The separation of bipolar (manic-depressive) from unipolar recurrent depressive psychoses. *Behav. Neuropsychiatry* 1:17–24
19. Biegel, A., Murphy, D. L. 1971. Unipolar and bipolar affective illness: Differences in clinical characteristics accompanying depression. *Arch. Gen. Psychiatry* 24:215–20
20. Buchsbaum, M., Goodwin, F. K., Murphy, D. L., Borge, G. 1971. Average evoked response in affective disorders. *Am. J. Psychiatry* 128:19–25
21. Buchsbaum, M., Landau, S., Mur-

phy, D. L., Goodwin, F. K. 1973. Average evoked response in bipolar and unipolar affective disorders: Relationship to sex, age of onset and monoamine oxidase. *Biol. Psychiatry* 7:199–212

22. Leonhard, K., Korff, I., Schulz, H. 1962. Die temperamente in den familien der monopolaren und bipolaren phasischen psychosen. *Psychiatr. Neurol.* 143:416–34
23. Reich, T., Clayton, P. J., Winokur, G. 1969. Family history studies. V. The genetics of mania. *Am. J. Psychiatry* 125:1358–70
24. Goetzl, V., Green, R., Whybrow, P., Jackson, R. 1974. X linkage revisited: A further family history study of manic-depressive illness. *Arch. Gen. Psychiatry* 31:665–72
25. Goodwin, F. K., Murphy, D. L., Bunney, W. E. Jr. 1969. Lithium carbonate treatment in depression and mania. *Arch. Gen. Psychiatry* 21:486–96
26. Goodwin, F. K., Murphy, D. L., Dunner, D. L., Bunney, W. E. Jr. 1972. Lithium response in unipolar vs. bipolar depression. *Am. J. Psychiatry* 129:44–47
27. Baron, M., Gershon E. S., Rudy, V., Jonas, Z., Buchsbaum, M. 1975. Clinical polarity, average evoked response, catechol-o-methyltransferase, and familial correlates of lithium response in depression. *Arch. Gen. Psychiatry* 32:1107–11
28. Mannarino, E., Kirshner, N., Nashold, B. S. 1963. The metabolism of (C^{14}) noradrenaline by cat brain in vivo. *J. Neurochem.* 10:373–79
29. Maas, J. W., Landis, D. H. 1968. In vivo studies of the metabolism of norepinephrine in the central nervous system. *J. Pharmacol. Exp. Ther.* 157:493–502
30. Sjoquist, B. 1975. Mass fragmentographic determination of 4-hydroxy-3-methoxymandelic acid in human urine, cerebrospinal fluid, brain and serum using a deuterium-labelled internal standard. *J. Neurochem.* 24:199–201
31. Karoum, F., Neff, N. H., Wyatt, R. J. 1976. Distribution and turnover rate of vanillylmandelic acid and 3-methoxy-4-hydroxyphenylglycol in rat brain. *J. Neurochem.* 27:33–35
32. Korf, J., Roth, R. H., Aghajanian, G. K. 1973. Alterations in turnover and endogenous levels of norepinephrine in the cerebral cortex following electrical stimulation and acute axotomy of cerebral noradrenergic pathways. *Eur. J. Pharmacol.* 23:276–82
33. Korf, J., Aghajanian, G. K., Roth, R. H. 1973. Stimulation and destruction of the locus coeruleus: Opposite effects on 3-methoxy-4-hydroxyphenylglycol sulfate levels in the rat cerebral cortex. *Eur. J. Pharmacol.* 21:281–89
34. Maas, J. W., Landis, D. H. 1971. The metabolism of circulating norepinephrine by human subjects. *J. Pharmacol. Exp. Ther.* 177:600–12
35. Ebert, M. H., Kopin, I. J. 1975. Differential labelling of origins of urinary catecholamine metabolites by dopamine-^{14}C. *Trans. Assoc. Am. Physicians* 88:256–64
36. Maas, J. W., Hattox, S. E., Martin, D. M., Roth, R. H. 1977. A direct method for studying brain amine metabolism in awake animals. In *Biochemistry and Function of Monoamine Enzymes*. New York: Dekker. In press
37. Maas, J. W., Fawcett, J., Dekirmenjian, H. 1968. 3-Methoxy-4-hydroxyphenylglycol (MHPG) excretion in depressive states: A pilot study. *Arch. Gen. Psychiatry* 19:129–34
38. Jones, F., Maas, J. W., Dekirmenjian, H., Sanchez, J. 1975. Diagnostic subgroups of affective disorders and their urinary excretion of catecholamine metabolites. *Am. J. Psychiatry* 132:1141–48
39. Taube, S. L., Kirstein, L. S., Sweeney, D. R., Heninger, G. R., Maas, J. W. 1977. Urinary 3-methoxy-4-hydroxyphenethyleneglycol (MHPG) and psychiatric diagnosis: Comparison of patients with primary affective disorders or schizophrenia and healthy comparison subjects. *Am. J. Psychiatry*. In press
40. Schildkraut, J. J., Keeler, B. A., Papousek, M., Hartmann, E. 1973. MHPG excretion in depressive disorders: Relation to clinical subtypes and desynchronized sleep. *Science* 181:762–64
41. Maas, J. W., Dekirmenjian, H., Jones, F. 1973. The identification of depressed patients who have a disorder of NE metabolism and/or disposition. In *Frontiers in Catecholamine Research,* eds. E.

Usdin, S. Snyder, pp. 1091–96. New York: Pergamon

42. Goodwin, F. K., Post, R. M. 1975. Studies of amine metabolites in affective illness and schizophrenia: A comparative analysis. In *Biology of the Major Psychoses: A Comparative Analysis*, ed. D. X. Freedman, pp. 249–332. New York: Raven. 372 pp.
43. Garfinkel, P. E., Warsh, J. J., Stancer, H. C., Godse, D. D. 1977. CNS monoamine metabolism in bipolar affective disorder. *Arch. Gen. Psychiatry* 34:735–39
44. Maas, J. W., Fawcett, J. A., Dekirmenjian, H. 1972. Catecholamine metabolism, depressive illness and drug response. *Arch. Gen. Psychiatry* 26:252–62
45. Schildkraut, J. J. 1973. Norepinephrine metabolites as biochemical criteria for classifying depressive disorders and predicting response to treatment: Preliminary findings. *Am. J. Psychiatry* 130:695–98
46. Beckmann, H., Goodwin, F. K. 1975. Antidepressant response to tricyclics and urinary MHPG in unipolar patients. *Arch. Gen. Psychiatry* 32:17–21
47. Fawcett, J., Maas, J. W., Dekirmenjian, H. 1972. Depression and MHPG excretion: Response to dextroamphetamine and tricyclic antidepressants. *Arch. Gen. Psychiatry* 26:246–51
48. Fawcett, J., Siomopoulos, V. 1971. Dextroamphetamine response as a possible predictor of improvement with tricyclic therapy in depression. *Arch. Gen. Psychiatry* 25:247–55
49. Maas, J. W. 1977. The effects of psychopharmacological agents on central nervous system amine metabolism in man. *Ann. Rev. Pharmacol. Toxicol.* 17:411–24
50. Garelis, E., Young, S. N., Lal, S., Sourkes, T. L. 1974. Monoamine metabolites in lumbar CSF: The question of their origin in relation to clinical studies. *Brain Res.* 79:1–8
51. Ashcroft, G. W., Crawford, T. B., Eccleston, D., Sharman, D. F. 1966. 5-Hydroxyindole compounds in the cerebrospinal fluid of patients with psychiatric or neurological diseases. *Lancet* 2:1049–52
52. Coppen, A. J., Prange, A. J. Jr., Whybrow, P. C., Noguera, R. 1972. Abnormalities of indoleamines in affective disorders Arch. Gen. Psychiatry 26:474–78
53. Mendels, J., Frazer, A., Fitzgerald, R. G., Ramsey, T. A., Stokes, J. W. 1972. Biogenic amine metabolites in cerebrospinal fluid of depressed and manic patients. *Science* 175:1380–82
54. Van Praag, H. M., Korf, J. 1971. Endogenous depressions with and without disturbances in the 5-hydroxytryptamine metabolism: A biochemical classification? *Psychopharmacologia* 19:148–52
55. Bowers, M. B., Heninger, G. R., Gerbode, F. 1969. Cerebrospinal fluid 5-hydroxyindoleacetic acid and homovanillic acid in psychiatric patients. *Int. J. Neuropharmacol.* 8:255–62
56. Goodwin, F. K., Post, R. M., Dunner, D. L., Gordon, E. K. 1973. Cerebrospinal fluid amine metabolites in affective illness: The probenecid technique. *Am. J. Psychiatry* 130:73–78
57. Asberg, M., Thoren, P., Traskman, L., Bertilsson, L., Ringberger, V. 1976. "Serotonin depression"—a biochemical subgroup within the affective disorders. *Science* 191: 478–80
58. Greenspan, J., Schildkraut, J. J., Gordon, E. K., Baer, L., Aronoff, M. S., Durell, J. Jr. 1970. Catecholamine metabolism in affective disorder. III: 3-methoxy-4-hydroxyphenylglycol and other catecholamine metabolites in patients treated with lithium carbonate. *J. Psychiatr. Res.* 7:171–83
59. Bond, P. A., Jenner, F. A., Sampson, G. A. 1972. Daily variations of the urine content of MHPG in two manic-depressive patients. *Psychol. Med.* 2:81–85
60. Jones, F., Maas, J. W., Dekirmenjian, H., Fawcett, J. A. 1973. Urinary catecholamine metabolites during behavioral changes in a patient with manic-depressive cycles. *Science* 179:300–2
61. Post, R. M., Stoddard, F. J., Gillin, J. C., Buchsbaum, M., Runkle, D. C., Black, K. E., Bunney, W. E. Jr. 1977. Alterations in motor activity, sleep and biochemistry in a cycling manic-depressive patient. *Arch. Gen. Psychiatry* 34:470–79
62. Sweeney, D. R., Maas, J. W. 1977. Stress and noradrenergic function in

depression. In *Psychobiology of the Depressive Disorders: Implications for the Effect of Stress,* ed. R. A. Depue. New York: Academic. In press

63. Sweeney, D. R., Maas, J. W., Heninger, G. R. 1977. State anxiety and urinary MHPG. *Psychosom. Med.* 39:57 (Abstr.)
64. Prange, A. 1973. The use of drugs in depression: Its theoretical and practical bases. *Psychiatr. Ann.* 3:56–75
65. Prange, A., Wilson, I., Lynn, C. W., Alltop, L. B., Stikeleather, R. A. 1974. L-Tryptophan in mania: Contribution to a permissive hypothesis of affective disorders. *Arch. Gen. Psychiatry* 30:56–62
66. Goodwin, F. K., Rubovits, R., Jimerson, D., Post, R. M. 1977. Serotonin and norepinephrine subgroups in depression. *Sci. Proc. Am. Psychiatr. Assoc.* 180:108 (Abstr.)

Ann. Rev. Med. 1978. 29:231–83
Copyright © 1978 by Annual Reviews Inc. All rights reserved

IMMUNOTHERAPY FOR MALIGNANT DISEASE

◆7279

James E. Goodnight, Jr., M.D., Ph.D. and Donald L. Morton, M.D.
Division of Oncology, Department of Surgery, University of California School of Medicine, Los Angeles, California 90024; and the Department of Surgical Services, Veterans Administration Hospital, Sepulveda, California 91343.

INTRODUCTION

Immunotherapy has been tested widely for the treatment of malignant disease because of its potential for a highly selective systemic attack on tumor cells. Unlike conventional treatment modalities that act directly on the cancer cell, immunotherapy works by activating host defenses to control and destroy tumor growth. Such an approach seems feasible, since host resistance appears to be involved in survival from cancer. A dramatic example of this phenomenon is the well-documented occurrence of spontaneous regression of human tumors (1). Moreover, it has been found that tumors possess unique antigens not present in normal tissues, and that these antigens elicit measurable immune responses in the tumor host (2–7). Various means for stimulating these responses exist, and these maneuvers clearly confer protection against tumor growth in experimental animals. Above all, investigators are reporting that stimulation of host defenses is beneficial for selected cancer patients. The hundreds of clinical trials recorded in the International Registry of Tumor Immunology (8) are a measure of the interest in immunotherapy for malignant disease.

In the early 1900s, success with immunization against infectious disease, plus the observation that tumor transplants in animals were rejected, led investigators to think that patients might be immunized against cancer. However, they failed to recognize that in outbred populations tumor transplants are rejected on the basis of normal histocompatibility antigens. Thus, resistance specific for tumors simply could not be demonstrated at that time. Nevertheless, attempts at immunotherapy for cancer were reported (9). After Coley observed prolonged regression of a neck sarcoma in a patient suffering from erysipelas, he tried injecting mixtures of live and heat-killed bacteria into other patients with various tumors (9). He reported some success with this treatment and the agents became known as Coley's toxins. The interest declined, however, when these experiments could not be reproduced.

0066-4219/78/0401-0231$01.00

Solid experimental evidence for host immunity to tumors appeared in the 1950s. In careful transplantation experiments with methylcholanthrene (MCA)–induced sarcomas in inbred mice, specific, tumor-associated antigens were identified (2, 3). Members of the same inbred strain are so nearly genetically identical that they normally do not reject each others' tissues. Foley (2) showed that when an MCA sarcoma was transplanted to a recipient syngeneic to the donor, allowed to grow and then excised completely, the recipient became resistant to further challenges with that tumor. In the same system, Prehn & Main (3) demonstrated that skin grafts from the tumor donor would neither induce tumor resistance in the recipient nor be rejected, even if the recipient had been immunized to the tumor. Subsequently, Klein (10) showed that these MCA sarcomas were immunogenic in their host of origin. These findings established the existence of tumor-specific transplantation antigens and were critical to the development of tumor immunology.

Interest in immunotherapy for malignant disease was rekindled in the late 1960s when three reports of successful clinical immunotherapy appeared. In a series of patients in remission from acute lymphoblastic leukemia, Mathé and co-workers (11) found that duration of remission and survival could be markedly extended by immunostimulation with Bacillus Calmette-Guerin (BCG) inoculation and/or vaccination with allogeneic leukemia cells. Morton and his colleagues (12) observed that 90% of cutaneous metastases of malignant melanoma would regress following intralesional injection with BCG. Moreover, in 17% of these patients, uninjected nodules regressed concomitantly with injected lesions. A third report by Klein (13) indicated that the majority of superficial basal cell and squamous cell carcinomas of the skin would resolve if delayed cutaneous hypersensitivity was induced in the area by repeated applications of various chemicals.

This successful use of immunotherapy for malignant disease generated a wave of enthusiasm. For example, in the 10 year period from 1965 to 1975, the volume of publications per year on tumor immunology and immunotherapy increased from less than 200 to 18,000 papers (Figure 1). It seemed that a new era of cancer treatment had begun. However, this enthusiasm ignored the reality of experimental tumor systems, which had shown that the potency of immunotherapy was limited. Animals specifically sensitized to resist the growth of a particular tumor are still susceptible if the tumor challenge is large enough (i.e. greater than 10^7 cells). It has likewise become apparent in clinical studies that immunotherapy used alone is largely ineffective in achieving regression of established visceral malignancies. The number of tumor cells in humans that can be controlled by host defense mechanisms is approximately 10^8 (14). Measurable tumors contain many more cells than this. Recognition of this limitation has tempered enthusiasm for immunotherapy, and a more realistic attitude now prevails.

The mechanisms underlying immunotherapy are complex and far from understood. Nevertheless, specific tumor immunity is a legitimate biological phenomenon. In the transplantation experiments cited, specific, tumor-associated antigens have been identified (2, 3). Both serological and cell-mediated immune responses to these antigens can be demonstrated (4), and animals can be specifically sensitized to resist the growth of a particular tumor.

In addition, there are nonspecific mechanisms of host resistance to cancer. Macrophages activated by endotoxin or other stimuli can selectively identify and destroy neoplastic cells based on some means of recognition other than antigenic differences (15). Macrophages also can be specifically armed by T-lymphocytes to attack tumor cells (16). A combined specific effector mechanism exists whereby antitumor antibody provides the specificity, and a cytotoxic effector cell with a receptor for the immunoglobulin molecule attaches and completes the destruction of the tumor cell (17). Some of these investigations extend to human tumor-host interactions. However, much remains to be learned, particularly to what extent these mechanisms explain the observed

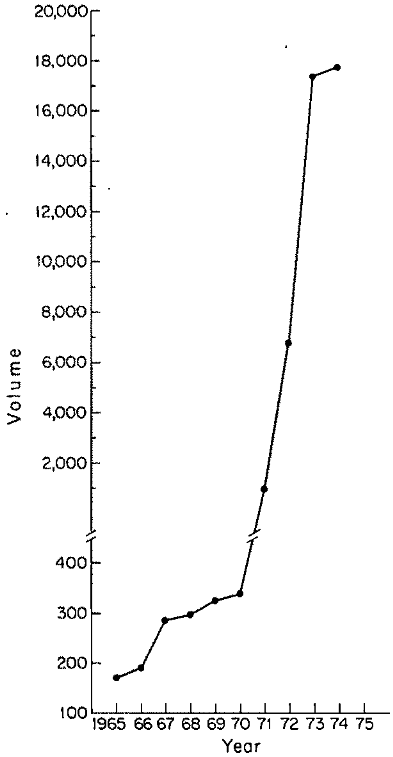


Figure 1. The increased incidence of literature from 1965 to 1975. Note the sharp rise from a little over 300 to 1000 from 1970–71 (MEDLARS, National Library of Medicine).

responses to immunotherapy. This subject is beyond the scope of this review, but it has been thoroughly discussed in several excellent reviews (18–22).

Despite the limited potency of immunotherapy and the lack of knowledge of its basic mechanisms, there is compelling evidence from clinical trials to indicate that immunotherapy benefits selected cancer patients. Enough clinical experience with immunotherapy has been acquired to allow for preliminary conclusions. The purpose of this review is to put immunotherapy for malignant disease in perspective based on a careful examination of the majority of published clinical trials.

ELEMENTS OF IMMUNOTHERAPY

Immunotherapy may be categorized as nonspecific, active-specific, or passive. For nonspecific immunotherapy, agents such as BCG or *Corynebacterium parvum*, both potent stimulators of host resistance, are used to trigger antitumor responses. Active-specific immunotherapy elicits antitumor immune responses by vaccination with tumor cells or portions thereof. These vaccines may be modified in various ways to enhance their immunogenicity. For passive or adoptive immunotherapy, immune effectors from a specifically sensitized host are transferred to the cancer patient in an effort to provide specific antitumor reactivity.

Nonspecific Immune Therapy

BCG BCG is an attentuated, live organism developed from a virulent strain of *Mycobacterium bovis* by Calmette and Guerin at the turn of the century (23). It has been used successfully throughout the world to immunize populations against tuberculosis. In addition, it is a potent stimulator of humoral and cellular immune responses (24). Most pertinent to this discussion is the fact that BCG inoculation of animals can induce resistance to tumor transplants and regression of some established tumors. This experimental work was recently reviewed by Bast and colleagues (18). Several factors are essential for successful BCG therapy in animal tumor models: (*a*) Tumor burden must be low. (*b*) The host must be immunocompetent. (*c*) Adequate amounts of BCG must be given. (*d*) BCG must be in close association with the tumor cells. (*e*) The tumor itself must be immunogenic. Similar observations have been made when BCG is used for human neoplasia.

The mechanisms underlying tumor regression following BCG injection are not completely understood. BCG does not appear to have a direct antitumor effect since tumor regression occurs only if the host raises an inflammatory response to the BCG. The most striking responses are seen following intralesional injection, which suggests that tumor destruction is a bystander effect of the hypersensitivity response to the BCG (12). However, Klein & Klein (25) performed an elegant experiment showing that immune responses are exquisitely specific for destruction of experimental tumors. They injected a mixture of tumor cells into an animal sensitized to only one of the tumor lines in the mixture. An inflammatory response resulted at the site of injection, and the line to which the animal was resistant was rejected. However, the other tumor line survived and grew out in the midst of the reaction.

There is no clear evidence that antibodies or specifically sensitized cells are responsible for tumor destruction following BCG injection, but specific antitumor responses do occur in this setting. It has been shown that the titer of antimelanoma antibody increases in melanoma patients treated with BCG (12). In a guinea pig hepatoma model, Hanna (26) also showed that antitumor antibody increased after BCG injection. Moreover, once the tumor regressed, the animal resisted late tumor challenges in sites distant from the original tumor, which indicates a systemic effect.

Side effects of BCG treatment are few when it is given by the Tine technique or with the Heaf gun, and usually consist of local inflammation, low-grade fever, malaise, and mild ulceration. With intralesional injection, more serious toxicities such as anaphylaxis, granulomatous hepatitis, systemic BCG infection, high fever and chills, or abscesses can occasionally occur (27). For these reasons, BCG should be administered only by personnel experienced in its use. BCG has been used widely in cancer therapy for local injection of accessible tumors, as adjuvant therapy for minimum residual disease, and in combination with chemotherapy for disseminated disease.

CORYNEBACTERIUM PARVUM *C. parvum* is a gram-positive anaerobic organism, which, in contrast to BCG, is killed by physical or chemical means when used for immunotherapy. In animals, treatment with *C. parvum* inhibits the growth of tumor inoculations, and, at least for MCA sarçomas in mice, can cause regression of established tumor (28–30). Given intralesionally, *C. parvum* induces regression of cutaneous malignancies. *C. parvum* is maximally effective when injected into the regional lymph node drainage of the tumor (32).

C. parvum is a consistent macrophage stimulator, which may be the basis of its antitumor activity, since it can produce inhibition of tumor growth in animals depleted of T cells (31). Some of its effects depend on the route of administration. Given intravenously, *C. parvum* suppresses T cells, but given subcutaneously, it stimulates T-cell activity. It is an active potentiator of antibody production in response to a variety of antigens (32).

In man, *C. parvum* has been tested primarily as an adjuvant to chemotherapy for advanced malignant disease (33). Side effects from *C. parvum* therapy such as fever, chills, nausea, vomiting, and changes in blood pressure, are seen mainly with intravenous administration (34). Local inflammation results from subcutaneous administration.

LEVAMISOLE Levamisole, a drug developed and used successfully as an antihelminthic, also acts to restore depressed immune responses. Renoux & Renoux (35) showed that mice treated with levamisole have increased resistance to *Brucella abortus* infection. In man, delayed cutaneous hypersensitivity responses are augmented by the drug (36). In addition, antibody responses to a variety of stimuli such as influenza are potentiated (37). Macrophage activity is increased after levamisole treatment as indicated by enhanced phagocytic activity (38).

Levamisole has been reported to be beneficial for several benign diseases. It helps clear aphthous stomatitis and warts. It gives symptomatic relief in systemic lupus, rheumatoid arthritis, and herpes simplex (32). Results of testing levamisole as an antitumor agent in animals have been inconsistent, although its use as a stimulant of host resistance suggests that it might be useful in treating malignancy. It has been tested in humans as an adjuvant to primary treatment for breast and lung carcinoma and initial results suggest benefit (39, 40). Side effects are low and are limited to mild GI distress and some fatigue, although some cases of agranulocytosis have been reported (41).

Active-Specific Immunotherapy

The primary goal of immunotherapy is to sharpen the specificity and augment the potency of immune responses that eradicate tumor cells. The most direct and specific way of achieving this goal is to stimulate the immune system with tumor antigens, in spite of their weak immunogenicity. The way in which an antigen is presented to the immune system influences the response. Altering the route of sensitization, the antigen dose, or the configuration of tumor antigens can result in a more effective antitumor response.

Alteration of the tumor cell surface may increase the immunogenicity of tumor antigens. The capacity of mouse leukemia cells to stimulate tumor transplantation resistance can be markedly increased by exposing the cells to neuraminidase, an enzyme obtained from *Vibrio cholera* (42). Neuraminidase cleaves sialic acid from cell surface glycoproteins, which may increase antigen exposure. Another method of increasing immunogenicity is to infect tumor cells with virus. The virus particles on the cell membrane may act as helper antigens, which in turn augment stimulation of the immune system (43). Lysates of virus-infected tumor cells have been used clinically to treat several kinds of cancer.

Intradermal injections of viable allogeneic tumor cells or tumor cell extracts are being used clinically to treat malignant melanoma, breast and lung cancer, soft-tissue sarcomas, and acute leukemia (44–47, 42). Animal experiments have shown that intradermal injection is a good route for sensitization. Moreover, since tumors of the same histologic type share common antigens, allogeneic tumor cells effectively stimulate tumor transplantation resistance (48). Because allogeneic cells are rejected on the basis of their normal histocompatibility antigens, the problem of tumor growth at the site of inoculation is eliminated. Clinical trials with viable allogeneic tumor vaccines have confirmed their safety and effectiveness. Immune responses can be augmented by mixing these tumor vaccines with immunostimulants such as complete Freund's adjuvant.

Passive Immunotherapy

Passive immunotherapy involves adoptive transfer of immune cells or antiserum with appropriate tumor specificity to the cancer patient. An alternative, theoretical approach is the transfer of subcellular fractions of immune cells such as immune RNA or transfer factor to induce specific antitumor reactivity in the patient's own lymphocytes. This approach is logical insofar as the agent transferred has activity against the tumor, and the recipient's immune system will accept and utilize it. However, logistical problems are involved in obtaining an adequate quantity of immunological agents with appropriate specificity and activity.

One precept of tumor immunology has been that of cell-mediated immunity as the primary effector function responsible for tumor destruction. For example, immunocompetent cells can be transferred from an inbred immunized animal to a nonimmunized animal to achieve tumor control. Control can be maintained as long as the cells persist. However, adoptively transferred allogeneic cells are eventually rejected, so that a continuous source of sensitized cells must be found. In humans, such a supply is usually unavailable, although one success has been achieved by transfer of blood from a patient who experienced a complete remission of melanoma (49). Another approach involves matching patients with the same tumor, sensitizing them to each other's tumor with subcutaneous transplants, and then exchanging their leukocytes. Response rates of 15–20% have been reported (50–52). This is at the price of some serious transfusion reactions, since allogeneic cells sensitize the patient to HL-A antigens. The use of immunocompetent cells from closely matched siblings is an alternative.

Passive transfer of antitumor antisera can induce tumor regression in some animal models (53). The antibody may be directly cytotoxic, or it may provide the specificity for cellular immune effectors with Fc receptors (17). Serotherapy is problematic though, because human sources of specific antitumor antisera for clinical use are often difficult to find, and heterologous antisera may cause serum sickness or anaphylaxis. Moreover, because antibodies occasionally can enhance tumor growth, there is some risk to this method. Serotherapy has been tried in patients with Burkitt's lymphoma, malignant melanoma, and renal cell carcinoma on a limited basis, but no systematic, controlled, clinical trial has been reported (53).

Certain informational molecules obtained from sensitized lymphocytes have been shown to confer specific reactivity on unsensitized lymphocytes. The virgin cells become sensitized without exposure to the antigen. One such informational agent is a low-molecular-weight, dialyzable substance known as transfer factor. Originally described by Lawrence a few years ago (54), this molecule transfers cell-mediated, but not humoral, responses to an antigen. Transfer factor has been used successfully to treat congenital immunodeficiency diseases such as Wiskott-Aldrich syndrome, ataxia-telangiectasia, and chronic mucocutaneous candidiasis. These advances led to limited clinical trials in cancer patients (55–57). A source of transfer factor with appropriate tumor specificity is a problem. Cohabitants of cancer patients have been used, and initial results suggest benefit, although these preliminary reports are largely anecdotal (57).

Mannick & Egdahl (58) showed that specific transplantation immunity could be transferred in rabbits using RNA extracted from sensitized lymphocytes. Pilch and co-workers (59) confirmed these observations, and extended them to antitumor responses. They showed that RNA extracted from lymphocytes of guinea pigs sensitized to a mouse tumor would confer specific antitumor reactivity on untreated mouse spleen cells. To clearly establish activity against specific tumor-associated antigens, they demonstrated the same phenomenon in a strain of inbred rats in which the tumor donor, effector cells, and RNA donor were syngeneic. This excluded any activity against normal histocompatibility antigens.

This demonstration of a xenogeneic immune RNA that would transfer specific reactivity was important for clinical trials, since it meant that appropriately sensitized animals could be used as a source of antihuman tumor-immune RNA. Immune RNA has virtually no toxicity, since lymphocytes cannot be instructed to attack "self"-antigens. Allergic reactions to the RNA are not a problem, since it is of low immunogenicity. Human lymphocytes exposed to immune RNA from sheep sensitized to a human tumor will destroy that tumor in vitro. Controls indicate that the activity is specific for the tumor and dependent upon exposure to the RNA. Initial clinical trials in humans with xenogeneic immune RNA recently were summarized by Ramming and colleagues (60).

CLASSIFICATION OF CLINICAL TRIALS

Although immunotherapy seems to be an important tool in the management of cancer, its full potential is unproven except for intralesional or topical treatment of accessible tumors. Its efficacy can be established only if published studies present an accurate and undistorted view. Results from clinical trials of immunotherapy vary considerably in their yield of information and reliability. The structure of clinical trials ranges from uncontrolled series to those in which patients are randomly assigned to treatment and control groups. Choice of controls, whether randomized, selected, historical, etc., affects the credibility of the data. Moreover, the conduct of clinical trials varies markedly in the number and selection of patients, the accuracy of diagnoses and staging, the period of follow-up, and the statistical treatment of results. The significance of a clinical trial depends on the clarity with which it excludes alternative explanations for its findings, which is in turn dependent upon the variables mentioned.

In this review, we attempt to classify clinical trials on the basis of how well each was able to show therapeutic benefit from immunotherapy for malignant disease. The trials are classified in order to evaluate and compare them as objectively as possible. In a review, the reader must necessarily view the subject through the eyes of the reviewer, which in turn introduces a major subjective influence. While a study may be highly interesting, it may have little in the way of substantive results. The conclusions presented in a particular study may reflect the bias of the reviewer, but the classification acts as a stimulus toward evaluating the studies in a uniform manner. The intent of the classification system is to give the reader some idea of the relative merit and significance of the clinical trials included in this review. The numbers assigned, listed in parenthesis after the trial, do not necessarily imply ranking, but rather serve to designate the category in which the study fell.

Type I studies were those that demonstrated that immunotherapy was an effective treatment for malignant disease on the basis of a frequent, clear-cut, and predictable response to the therapy. Studies in this category are mainly those limited to intralesional injection or topical treatment of accessible tumors. Confirmed studies were designated Type A, and unconfirmed studies were designated Type B. Additional criteria for inclusion in this category were: (*a*) Five or more patients must have shown clearly objective responses (greater than 50% reduction of tumor, lasting longer than one month). (*b*) The five or more responders must have represented 50% or more of the patients treated.

Type II studies were those that demonstrated therapeutic benefit based on a prospective, randomized, controlled trial. These studies had to have a definite therapeutic objective, an adequate number of patients, well-defined diagnoses and staging, and statistically significant results. In clinical trials, it is very difficult to control the myriad variables that influence results, and there may be unsuspected factors that introduce unintentional bias in the work. Random assignment of patients to treatment and control groups helps to balance these variables and ensure that prognostic factors are equal in the two groups.

Type III clinical trials were controlled but the assignment to treatment and control groups was not randomized. The key requirement for this category was the inclusion of data in the published results that allowed us to judge whether prognostic factors were equal in the treatment and control groups. With this proviso, studies with either simultaneous or sequential control groups were included in this category. Similar to Type II studies, the Type III studies had to have definite diagnoses and staging, and statistically valid results. The main evidence for therapeutic benefit from immunotherapy has come from Type I, II, or III studies. Only 20% of the studies reviewed fell into these classifications. The other classifications to be described suggest areas where immunotherapy may be beneficial, but they do not establish its value for treatment of malignant disease.

Type IV studies were controlled trials that demonstrated therapeutic benefit but the control group could not be fully evaluated from the data given. Most studies with historical controls fit into this category. It was difficult to judge whether prognostic factors were comparable in the earlier treatment groups.

Subtle changes in overall management can occur, besides the treatment being tested. Studies with simultaneous controls could be included here also, depending upon our assessment of the control group. Studies that compared treatment results to published series from other institutions were not considered controlled trials.

Type V studies were uncontrolled clinical trials that showed some therapeutic benefit (either complete tumor regressions or significant partial regressions), more than would be expected from spontaneous improvement. Studies were included in this category if two or more patients showed clearly objective responses (greater than 50% reduction in tumor, lasting longer than one month), and this result represented 15% or more of the patients tested.

Type VI studies were clinical trials in which therapeutic benefit could not be evaluated from the data given. This category included many studies that failed to show therapeutic benefit but were not designed to yield significant negative information. Examples include uncontrolled studies on small numbers of patients who had objective responses so infrequently as to be unevaluable. Controlled trials with either positive or negative results could be classified here for reasons such as an inappropriate control group, too few patients, or results that were not statistically significant. Examples include studies that were inadequately controlled for treatments other than the immunotherapy. Another group included here was uncontrolled trials testing adjuvant immunotherapy for control of microscopic residual disease. Since there was no measurable tumor, the end point was survival or duration of remission, and benefit in this situation cannot be evaluated in the absence of controls. In addition, several preliminary reports from well designed and well controlled studies were classified as Type VI if the data were inconclusive.

There were studies that clearly demonstrated that the method of immunotherapy being tested was of no therapeutic benefit. The criteria for evaluation of this group were essentially the same as for Type II studies except, in this instance, each produced significant negative results. These are designated as Type VII.

Our classification system is intended as a general guide. Obviously, the classifications are arbitrary, and not all readers would reach the same conclusions. However, every attempt was made to apply the criteria uniformly and to present a fair assessment of clinical investigation of immunotherapy for malignant disease.

CLINICAL TRIALS

Results from a large number of clinical trials of immunotherapy for malignant disease have been published. The effectiveness of this modality can best be appreciated by examination of the results obtained from treatment of the individual diseases. For this reason, the discussion of clinical trials has been arranged by site of disease.

Brain Tumors

The section on brain tumors (Table 1) is placed first in the discussion of clinical trials of immunotherapy for malignant disease strictly on an alphabetical basis. A priori, one would expect that immunotherapy for brain tumors is unlikely to be of benefit since the central nervous system seems to be an immunologically privileged site. Clinical trials, so far, have not invalidated this hypothesis. In 1969, Trouillas & Lapras (61) reported on the intrathecal injection of autologous, sensitized lymphocytes for control of glioblastoma multiforme (VI). The patients had pieces of tumor, removed at surgery, implanted subcutaneously. Their thoracic duct lymphocytes were then harvested and given intrathecally. Five patients were treated but no therapeutic benefit was demonstrated. Bloom and his colleagues (62) have been doing investigations of immunotherapy for glioblastoma since 1960. Their most recent report involved a randomized, prospective trial of 62 patients (VII). Both groups had surgical removal of tumor and irradiation. The immunotherapy patients were given intradermal, autologous tumor cells obtained from the surgical specimen. No therapeutic benefit was evident in patients receiving immunotherapy. All patients who received tumor cells were dead at 30 months, while of the 35 nonimmunotherapy patients, seven remained alive at that interval. Six of the seven patients died subsequently, and survival in the two groups showed no significant statistical difference. Thus far, there is no evidence of benefit from immunotherapy for malignant brain tumors.

Breast Carcinoma

In breast cancer, late recurrences of rapidly progressive disease are seen many years after primary treatment, which suggests that the disease was controlled by a suppressive mechanism in the interim. Moreover, the induction of local, delayed cutaneous hypersensitivity responses (DCH) in a chest wall recurrence of breast carcinoma, may cause it to regress. Such observations suggest that immunotherapy may be useful in the treatment of breast carcinoma. In addition to local control of skin metastases, immunotherapy has been used as an adjuvant to other treatments for control of both minimal residual disease and disseminated disease, as indicated in Table 2.

Intralesional injection of BCG has proved to be an effective means of control

Table 1 Brain tumors

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
1969	Trouillas & Lapras	61	Intrathecal autologous sensitized lymphocytes	VI
1973	Bloom et al	62	Intradermal autologous tumor cells	VII

for skin metastases from breast carcinoma. Complete regression of tumor was observed in seven of eight patients by Smith and colleagues (63) (I-A), and in seven of 14 patients by Klein (64) (I-A). In all instances, the responders had or developed positive PPD skin tests. While this treatment produced a striking effect on the local disease, the response was limited to the treated lesion, and no determination of whether the treatment extended survival was reported. One report implied failure of intralesional BCG in this setting with only three of 20 patients showing a response (VI) (65). Whether this failure had to do with variations in the technique or the dose of BCG was not clear. Both of these variables plus the level of tumor burden are critical to the success of immunotherapy.

Other agents that induce DCH responses have been used successfully in the local treatment of chest wall recurrences of breast cancer. Nitrogen mustard injected intralesionally caused complete regression of treated lesions in five of six patients sensitized to the drug (I-B) (66). Tumor regression apparently was not a direct effect of the drug since unsensitized patients were not affected, and only 75% of patients could be sensitized. In another study, topical application of dinitrochlorobenzene (DNCB) to skin metastases in DNCB-sensitized patients caused complete regression of lesions in three of eight patients tested (V) (67). Intralesional injection of autologous lymphocytes stimulated with phytohemagglutinin (PHA) caused one complete regression and two partial responses in

Table 2 Breast carcinoma

Year	Investigator	Ref.	Immunotherapy	Type of study
Local treatment for skin metastases				
1971	Stjernsward & Levin	67	Topical DNCB	V
1972	Cheema & Hersh	68	Intralesional PHA-activated lymphocytes	V
1973	Smith et al	63	Intralesional BCG	I-A
1975	Garas et al	65	Intralesional BCG	VI
1975	Goldman	66	Intralesional nitrogen mustard	I-B
1976	Klein et al	64	Intralesional BCG	I-A
Adjuvant therapy for minimum residual disease				
1974	Anderson et al	70	Surgery—tumor autotransplant	VI
1976	Rojas et al	39	Radiation—levamisole	III
Treatment for disseminated disease				
1974	Oettgen et al	74	Transfer factor	VI
1975	Israel & Edelstein	33	*C. parvum*—chemotherapy	II
1976	Gutterman et al	73	Intradermal BCG—chemotherapy	III
1977	Pinsky et al	72	*C. parvum*—chemotherapy	VI

skin metastases of breast carcinoma (V) (68). In this study, injection of adjacent lesions with unstimulated lymphocytes or saline had no effect.

The importance of these observations of successful local immunotherapy is twofold. First, immunotherapy offers a useful alternative for obtaining local control of accessible, recurrent breast cancer; and second, its potential for stimulation of host resistance may improve systemic control of tumor.

As an adjuvant to other treatment, immunotherapy has been used to prevent or delay recurrence following primary treatment for breast cancer. Rather striking results were obtained in a series of patients with Stage III disease (UICC classification) treated with radiation therapy followed by immunotherapy with levamisole (III) (39). Forty-three patients were irradiated (4000 R), and then alternately assigned to levamisole (150 mg by mouth daily for three days every two weeks) or a placebo. Minimum follow-up was 21 months. In the levamisole groups, median time to recurrence was 25 months vs nine months for those taking a placebo. Patient survival rate in the levamisole group was 90% at 30 months, whereas only 35% of the placebo group were alive at this interval ($P < .01$). Control and treatment groups were apparently equal for age, menopausal status, child bearing, and extent of disease. However, the control group reported here did rather poorly compared to an overall 70% survival rate at 30 months for patients with Stage III disease reported by the American Joint Committee for Cancer Staging and End Results Reporting (AJCCS) (69). Survival rate in the placebo group was half that expected from the AJCCS statistics, which suggests that factors affecting prognosis were not equal in the control and treatment groups. However, the levamisole group survival rate was better than the AJCCS figures, indicating possible therapeutic benefit. The results of this study have not been confirmed, but a prospective randomized trial is in progress (39).

An adjuvant study in which irradiated, autologous, tumor implants were used following radical mastectomy for breast carcinoma has been reported (VI) (70). No therapeutic benefit could be established from this attempt to stimulate specific antitumor responses.

Adjuvant chemotherapy following mastectomy has been shown to reduce the recurrence of locally advanced breast carcinoma (71). The addition of immunostimulation to this regimen might give even better results. Studies are in progress to test this hypothesis (46).

The combination of immunotherapy and chemotherapy may improve treatment results for disseminated breast carcinoma. In a prospective, randomized trial, Israel & Edelstein (33) used subcutaneous *C. parvum* in addition to five-drug chemotherapy for 43 patients with metastatic breast carcinoma (II). A control group of 39 patients received only the chemotherapy. A statistically significant improvement in survival time was seen in the patients receiving combination chemo-immunotherapy with a median survival time of over 18 months, compared to less than 12 months for the control group ($P < .001$). More recently, Pinsky and colleagues (72) combined subcutaneous *C. parvum* with a four-drug regimen for disseminated breast carcinoma. Preliminary results from their randomized trial indicate that responders in the chemo-immunotherapy group survived significantly longer than the responders on chemotherapy alone (VI). However, the survival curves for the two groups as a whole are not significantly different. Final evaluation must await completion of the study.

In another study of disseminated breast carcinoma, Gutterman (73) compared the use of repeated intradermal BCG combined with three-drug chemotherapy to the use of chemotherapy alone (III). Forty-five patients in the chemo-immunotherapy group were compared to 44 historical controls. In the latter group, 23 patients died with a median survival time for the group of 14 months. Although the chemo-immunotherapy group was followed for only 12 months, just five patients had died, giving this group a projected median survival time of much longer than 14 months. Although this preliminary result appears significant, improvements in overall treatment might have influenced the results. Nevertheless, the test and control groups were comparable for age, menopausal status, disease-free interval, and prior therapy. All patients with disseminated disease were entered consecutively and treated in the same institution.

Results from these chemo-immunotherapy trials are promising, and indicate that potent drugs and immunostimulation may be effective in combination. The contribution of immunotherapy is dependent upon the capacity of the chemotherapy to reduce tumor burden significantly. For breast cancer, this is feasible and additional prospective, randomized trials to confirm these findings are needed.

Transfer factor has been tested as a method for treating disseminated breast carcinoma (VI) (74). After other treatments had failed, six patients were given transfer factor obtained from lymphocytes of healthy women. Only one patient responded. However, this was a patient with inflammatory breast carcinoma who showed an arrest of progression of disease for six months in the absence of any other treatment. Although no conclusions can be drawn from a single result, this response was quite interesting.

It is clear that conventional modalities are still the mainstay of treatment for breast carcinoma. However, it appears that immunotherapy may be of benefit at several points in the treatment of malignant breast disease, and its use merits additional study.

Colo-Rectal Carcinoma and Other Gastrointestinal Tumors

Survival figures for colo-rectal carcinoma treated by surgery alone have remained the same for many years. For a patient with Dukes' C colo-rectal carcinoma, the five-year survival expectancy with surgery alone is an average 32% (75). In an effort to improve these figures, primary surgical treatment has been combined with adjuvant irradiation and/or chemotherapy. An interim report of a clinical trial of adjuvant immunotherapy following surgery for Dukes' C colo-rectal carcinoma has appeared (IV) (76). The adjuvant therapy group (83 patients) was randomized to receive long-term 5-Fluorouracil (5-FU) plus intradermal BCG (50 patients), or long-term intradermal BCG only (33 patients).

These groups were compared to historical controls treated by surgery alone at the M. D. Anderson Hospital and Tumor Institute in Houston, Texas (73 patients). For both adjuvant groups, statistically significant increases in the disease-free interval and survival rate were noted over controls at a median follow-up of approximately 16 months. There was no significant difference between the two adjuvant groups. This was a preliminary report, and a much longer follow-up period will be required to determine if the actual incidence of recurrence is decreased or the absolute number of survivors increased. Because adjuvant therapy of any kind is unproven in colo-rectal disease, this study would be more convincing if a surgery-only control group was included in the prospective randomization.

According to Moertel and colleagues (77), the addition of immunotherapy to chemotherapy for advanced colo-rectal carcinoma has not increased the expected 20% objective response rate (VI). Methanol extractable residue of BCG (MER) was used for the immunotherapy in this study. In an earlier study, Moertel (78) noted response to MER injections alone in three of 36 patients with advanced GI malignancy (VI), but could not confirm the result in a subsequent study (VI) (79). In a trial reported by Engstrom and co-workers (80), the combination of 5-FU and BCG was compared to 5-FU alone for the treatment of advanced colo-rectal carcinoma (VI). Again, there was no difference in the objective response rates. So far, immunotherapy does not appear to be of benefit in this situation.

Griffen & Meeker (81) used passive immunotherapy to treat 32 patients who had advanced colo-rectal carcinoma (V). Patients were paired for cross transplantation of their tumors. The patients then exchanged leukocytes. Five objective responses (16%) were reported, which is a response rate consistent with this type of treatment for other tumors. No follow-up report has appeared.

Falk and colleagues (82) used intraperitoneally injected BCG plus oral BCG combined with chemotherapy for the treatment of advanced malignant tumors of the GI tract (VI). Results of this treatment are difficult to evaluate as the study was uncontrolled and variable treatment regimens were used. However, for patients with colon carcinoma metastatic to the liver, whose short life expectancy as a group has been well established, some benefit may have been gained from the treatment. For nine patients treated, median survival time was 12 months, and two patients were alive at 19 and 24 months. This protocol should be tested in a controlled trial.

The use of thymic humoral factor in GI malignancies has been reported by Turowski and colleagues (83). This study was uncontrolled, and, from the data given, therapeutic benefit could not be evaluated.

As indicated in Table 3, clear-cut benefit from immunotherapy for GI tumors remains to be demonstrated. Chemotherapy is relatively impotent for these tumors and immunotherapy has not been additive. Results from surgical treatment plus adjuvant immunotherapy are still very preliminary. Considerable improvement is needed in all phases of treatment for gastrointestinal malignancies.

Genitourinary Tumors

Intradermal BCG was given to a group of patients with disseminated carcinoma of the prostate in an effort to augment host resistance (VI) (84). No significant clinical benefit could be identified. Merrin and colleagues (85) reported the use of intratumoral injections of BCG in patients with advanced carcinoma of the prostate (VI). In some patients, this treatment did produce necrosis in the injected tumor, but no alteration in clinical course was noted. In both studies, the patients had large tumor burdens, which diminished the possibility for successful immunotherapy (Table 4).

For the treatment of recurrent superficial bladder cancer, Eidinger & Morales (86) tried intracavitary instillations of BCG (VI). Although this was a preliminary report, their data suggested a decrease in recurrence rate. This was a reasonable trial since local immunotherapy is effective for other types of accessible tumors.

Xenogeneic immune RNA has been used in the treatment of disseminated renal cell carcinoma (VI) (87). The immune RNA was obtained from sheep sensitized to human hypernephromas. No tumor regressions were seen in 20 patients treated, but seven patients showed an apparent arrest of progression of disease lasting a minimum of three months. A subgroup of eight patients, with pulmonary metastases only, seemed to be surviving longer that a similar group of historical controls, but the follow-up was short and the numbers small. Nine patients with advanced local disease removed at surgery were treated with immune RNA (87, 88). Each was at high risk for recurrence, but no recurrence was observed over a mean follow-up of 18 months. Therapeutic benefit in this setting must be demonstrated in a controlled trial, however.

Table 3 Colo-rectal carcinoma and GI tumors

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
Adjuvant immunotherapy				
1976	Mavligit et al	76	BCG & 5-FU, Duke's C Colo-rectal cancer	VI
Mixed studies in advanced tumors				
1972	Griffen & Meeker	81	Tumor cell vaccine in advanced colon carcinoma	VI
1975	Moertel et al	78	MER injections	VI
1976	Falk et al	82	Oral and IP BCG in advanced GI tumors	VI
1976	Turowski et al	83	Thymic humoral factor in advanced GI tumors	VI
1976	O'Connell et al	79	MER injections	VI
1977	Engstrom et al	80	BCG, chemotherapy	VI
1977	Moertel et al	77	MER, BCG, chemotherapy	VI

Table 4 Genitourinary tumors

Year	Investigator(s)	Ref.	Carcinoma	Immunotherapy	Type of study
1973	Guinan et al	84	Metastatic prostate	Intratumor BCG	VI
1975	Merrin et al	85	Prostate	Intratumor BCG	VI
1976	Eidinger & Morales	86	Bladder	Intracavitary BCG	VI
1976	Skinner et al	88	Renal cell	Xenogeneic immune RNA	VI
1977	Ramming & deKernion	87	Renal cell	Xenogeneic immune RNA	VI

The role of immunotherapy for genitourinary tumors is as yet undetermined, but the studies identify important areas for further investigation.

Gynecological Cancers

A very interesting use of topical immunotherapy has been reported in post-hysterectomy patients with vaginal smears positive for malignant cells (I-B) (89). Positive cytological findings were documented on three repeat smears, then six patients with these findings were sensitized to DNCB. Once good DCH responses were obtained, an appropriate concentration of DNCB cream was carefully applied intravaginally, which produced an intense local reaction. Six weeks after treatment, repeat smears were all normal. In follow-up from 2 to 35 months (median, 21 months), all patients remained normal. The mechanism of response in these instances must be related to that for topical or intralesional immunotherapy of skin cancers and metastases. If the beneficial effects of the treatment can be confirmed in a larger number of patients, it should become a very useful therapy.

Immunotherapy by itself, or in combination with chemotherapy, has been tried for advanced gynecological cancer, but no clear evidence of effectiveness has appeared (Table 5). Tumor cell vaccines alone were apparently ineffective against established disease in one large study (VII) (90). A beneficial effect for advanced ovarian carcinoma using tumor cell vaccine combined with BCG has been claimed, but this result was based on one case with a short follow-up period (VI) (91). Uncontrolled trials combining chemotherapy and immunotherapy for various advanced gynecological tumors have shown some limited responses, but without control patients the contribution of immunotherapy cannot be determined (VI) (92, 93). Compared to patients treated with chemotherapy alone, patients with advanced ovarian carcinoma who received a combination of BCG, tumor cell vaccine, and chemotherapy had a significantly extended survival time, as reported by Hudson and colleagues (VI) (94). However, this was a small study based on historical controls with other variables besides immunotherapy existing between the test and control groups. These problems forestall any definitive conclusions, but this combination should be tested in a larger, more rigidly controlled trial.

Leukemia

ACUTE LYMPHOBLASTIC LEUKEMIA A major stimulus to interest in immunotherapy for malignant disease came from an early report by Mathé and coworkers (11), which showed that immunotherapy markedly prolonged remission and survival times in acute lymphoblastic leukemia (ALL) (II). Acute leukemia in remission was chosen as an appropriate clinical test of immunotherapy, because the low tumor burden offered a greater chance of success. Thirty patients with ALL in prolonged remission on chemotherapy had their drugs stopped. Ten of these patients were chosen randomly to receive no additional treatment, and all relapsed within five months. In contrast, 10 of 20 patients who were given immunotherapy that consisted of BCG and/or allogeneic leukemia cells remained in remission for more than 295 days. Seven of these patients have remained in remission for 7–12 years. Cures have probably been achieved since the treatment was stopped after five years (95).

Unfortunately, the usefulness of immunotherapy for ALL has not been established by subsequent studies, as seen in Table 6. One recent report by Otten (96) indicated that immunotherapy, consisting of BCG and allogeneic leukemia cells, was at least equal to chemotherapy in maintaining remission in ALL (II). Mathé and his group (97, 95) have continued to use immunotherapy for ALL and report a beneficial therapeutic effect. However, since the original report, they have not included a nonimmunotherapy control group in their studies. For this reason, and because chemotherapy is used whenever necessary to reinduce

Table 5 Gynecological cancers

Year	Investigator(s)	Ref.	Tumor type	Immunotherapy	Type of study
1962	Graham & Graham	90	Advanced gynecological tumors	Tumor cell vaccine	VII
1974	Imperato et al	91	Advanced ovarian carcinoma	Tumor cell vaccine and BCG	VI
1975	Guthrie & Way	89	Cytological vaginal carcinoma	Topical DNCB	I-B
1976	Pattillo	92	Advanced gynecological tumors	Tumor antigen vaccine, BCG, and chemotherapy	VI
1976	Hudson et al	94	Advanced ovarian carcinoma	Tumor cell vaccine, BCG, and chemotherapy	VI
1977	Rao et al	93	Advanced ovarian carcinoma	*C. parvum* chemotherapy	VI

remission in these studies, it is not possible to judge the therapeutic benefit gained from immunotherapy.

Two large, randomized, controlled studies have failed to show that BCG therapy is useful for maintaining chemotherapy-induced remissions of ALL (VII) (98, 99). Both studies compared repeated intradermal BCG treatment to chemotherapy for maintaining remission, and the latter treatment gave significantly better results. Results from BCG therapy were similar to no additional treatment. None of these studies reproduced the conditions of Mathé's original trial or that of Otten, because their immunotherapy consisted of BCG only and did not include allogeneic leukemia cells. From Mathé's view, treatments with BCG and BCG combined with leukemia cells were of equal efficacy. However, it may be that the contribution of specific-active stimulation by the leukemia cells is critical. In any case, chemotherapy for ALL has steadily improved, so it is unlikely that the conditions in Mathé's study will be reproduced (100). A recent report from the National Cancer Institute in which BCG and allogeneic leukemia cells were used in combination with current chemotherapy showed no advantage over the use of chemotherapy alone (VII) (101). There have been no other studies establishing therapeutic benefit from immunotherapy in ALL. They either have shown negative results (VI) (102, 103), or were too limited for definitive conclusions (VI) (104). At the present time, chemotherapy is the treatment of choice for ALL, and the role of immunotherapy is uncertain.

ACUTE MYELOGENOUS LEUKEMIA Immunotherapy for the treatment of acute myelogenous leukemia (AML) has been more promising (Table 7). In a prospective randomized trial of patients in remission, Powles (105) reported that the

Table 6 Acute lymphocytic leukemia

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
1968	Albo et al	125	Parental plasma, chemotherapy	VI
1969	Skurkovich et al	126	Reinfusion autologous cells and serum	IV
1969	Mathé et al	11	BCG, allogeneic leukemia cells	II
1971	Medical Research Council	98	BCG	VII
1972	Mathé et al	97	BCG, allogeneic leukemia cells	VI
1973	Leventhal et al	102	BCG, allogeneic leukemia cells, MTX, allogeneic leukemia cells	VI
1975	Ekert & Jose	104	BCG, chemotherapy	VI
1975	Heyn et al	99	BCG	VII
1975	Poplack et al	101	BCG, cells, chemotherapy	VII
1975	Sacks et al	103	RAJI cells, chemotherapy	VI
1976	Mathé et al	95	BCG, allogeneic leukemia cells	VI
1977	Otten	96	BCG, leukemia cells	II
1977	Kay	104a	BCG	VI

addition of immunotherapy with BCG and allogeneic leukemia cells to maintenance chemotherapy significantly prolonged survival time (II). Twenty-three patients receiving combined chemo-immunotherapy showed a median survival of 545 days compared to 303 days for patients receiving chemotherapy only ($P = .003$). Relapse rates and duration of remission were improved in the immunotherapy group, but were not significantly different from controls. In another prospective, randomized trial reported by Vogler & Chan (106), 41 patients with AML were given late consolidation chemotherapy, then randomly assigned to receive maintenance chemotherapy plus BCG or chemotherapy alone (II). There was a marked improvement in the duration of remission for 18 patients receiving chemo-immunotherapy: 39 weeks vs 26 weeks for the 23 control patients receiving drugs alone ($P = .002$).

Investigators from M. D. Anderson Hospital and Tumor Institute reported a trial of BCG therapy combined with chemotherapy for maintaining remission in patients with acute leukemia (IV) (107). The statistical treatment of the data and the controls in their initial report were criticized (108). However, they issued a revised report that maintained that there was improvement in the duration of remission for 14 AML patients treated with chemo-immunotherapy compared to 21 historical controls receiving chemotherapy only ($P = .04$) (109). Median survival time of those on chemotherapy alone was 78 weeks; for the group

Table 7 Acute myelogenous leukemia

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
1973	Powles et al	105	BCG, leukemia cells, chemotherapy	II
1973	Freeman et al	114	BCG, leukemia cells	VI
1974	Vogler & Chan	106	BCG, chemotherapy	II
1974	Gutterman et al	107	BCG, chemotherapy	IV
1976	Powles	115	BCG, leukemia cells	VI
1976	Whiteside et al	110	BCG, leukemia cells, chemotherapy	VI
1976	Holland & Bekesi	42	Neuraminidase-treated cells, chemotherapy	VI
1977	Reizenstein et al	111	BCG, leukemia cells, chemotherapy	II
1977	Peto	112	BCG, leukemia cells, chemotherapy	VI
1977	Sauter et al	122	Viral oncolysate, chemotherapy	VII
1977	Whittaker & Slater	117	i.v. BCG, chemotherapy	VI
1977	Cuttner et al	118	MER, chemotherapy	VI
1977	Gee & Clarkson	120	Heptavalent pseudomonas vaccine, chemotherapy	VI
1977	McIntyre et al	121	Poly I:Poly C, chemotherapy	VII

receiving combined therapy, a median survival time had not been reached, but it appeared that it was going to be more than 98 weeks. In a small, uncontrolled trial, Whiteside and co-workers (110) indicated a marked improvement in survival time and duration of remission for patients receiving chemotherapy plus BCG and allogeneic cells (VI). Using similar chemo-immunotherapy in a prospective, randomized trial, Reizenstein and colleagues (111) from Sweden reported a doubling of median survival time from 10 to 20 months over chemotherapy alone ($P < .05$) (II). Thus, several studies indicate that immunotherapy used in combination with chemotherapy has improved the results of treatment for AML patients who can be brought into remission initially with drugs.

There have been two negative reports on the use of combined chemo-immunotherapy in maintaining remission in patients with AML. Peto (112), reporting for the British Medical Council, notes that BCG plus allogeneic blast cells do not add to remission or survival times over chemotherapy used alone (VI). Hewlitt and colleagues (113) used BCG in addition to known potent chemotherapy, and found that there was a slight trend in favor of the combined therapy, but it was not statistically significant (VII). Confirmation of therapeutic benefit from chemo-immunotherapy in AML must await additional trials, and is not yet the standard of care for this disease.

The use of BCG and allogeneic leukemia cells without chemotherapy as maintenance treatment for AML was reported initially by Freeman and co-workers (114) (VI). They also did frequent bone marrow studies, and with any increase in blast cells they started reinduction chemotherapy promptly. It was their impression that second and subsequent remissions were much easier to obtain than previously. In a later report, Powles (115) indicated that immunotherapy given without chemotherapy for maintenance gave similar results to the combined treatment in prolonging survival time (VI). Both of these reports were from uncontrolled trials however. M. D. Anderson Hospital and Tumor Institute has reported the use of BCG alone to maintain patients with adult acute leukemia in remission after the cessation of chemotherapy (VI) (73). Thirty-seven leukemia patients in remission on long-term drugs were given late, intensive, consolidation chemotherapy, and then the drugs were stopped. Thirty patients received BCG as ongoing treatment; seven received no additional treatment. Superior survival time and duration of remission were indicated for the BCG group, though, the result was compared to a group of historical controls who had not received the same chemotherapy. In fact, the same group of historical controls was used for comparison to show the benefit of late consolidation chemotherapy in another publication (VI) (116). The BCG data also compared favorably to the seven patients not receiving immunotherapy, but this was an extremely small group for comparison. Thus, the benefit of immunotherapy used alone in this setting remains unproven.

Holland & Bekesi (42) tried an innovative approach to immunotherapy for adult patients with AML (VI). The patients were brought into chemotherapy-induced remission, and then were given maintenance chemotherapy plus neuraminidase-treated allogeneic leukemia cells. A very small initial clinical trial suggested that this treatment enhanced the duration of remission and survival in patients with AML. A larger, prospective, randomized trial is in progress in which 24 patients have received maintenance drugs only, while a second group of 27 patients have been given maintenance drugs plus neuraminidase-treated leukemia cells (J. G. Bekesi, personal communication). In the chemotherapy-only group, 22 patients (92%) relapsed with a median duration of 33 weeks. In contrast, the combined chemo-immunotherapy group of 19 patients (70%) are still in remission, 11 for more than 75 weeks and five for more than three years. This appears very promising, but evaluation of therapeutic benefit must await completion and publication of the study.

Other immunotherapy treatments for AML have been tried with variable success. BCG given intravenously in combination with maintenance chemotherapy for AML is reported to have an acceptable toxicity, and to yield a better survival than chemotherapy alone (VI) (117). MER combined with chemotherapy is being tested for AML, but results are preliminary (VI) (118). Schwarzenberg (119) reported some years ago that adoptive transfer of leukocytes to AML patients failing on other treatments would induce complete remission, although these remissions were short-lived (V). Other treatments used in combination with chemotherapy, including virally treated leukemia cells, pseudomonas vaccine, and polynucleotides, have not shown therapeutic benefit (VII) (120–122). Thymic humoral factor has been used in the treatment of various hematologic malignancies, and is reported to give symptomatic improvement, but the therapeutic benefit could not be evaluated from the data given (VI) (123).

OTHER LEUKEMIA STUDIES Clinical trials of immunotherapy for leukemia that employed passive transfer of immune serum have been reported (Table 8). DeCarvalho (124) immunized horses (and a donkey) against human tumor and then used γ-globulin extracted from the equine serum to treat tumor patients (V). Fifteen patients with leukemia or lymphoma who had failed on other therapies were treated. He reported seven complete responses to the treatment, including two patients who had remissions of more than 12 and 29 months, respectively. This was a preliminary study, and no follow-up or confirmation has been published. Surprisingly, he reported relatively few side effects from this xenogeneic serum. In treating children with ALL, Albo and colleagues (125) used plasma from parents in addition to maintenance chemotherapy (VI), but there was no beneficial therapeutic response. Skurkovich and co-workers (126) reported that in children being treated with chemotherapy for ALL reinfusion of autologous cells and serum obtained early in remission and stored increased the duration of remission (IV). This result was in comparison to selected control patients receiving chemotherapy only. This was a small group of patients, and there has been no follow-up or confirmation.

Immunotherapy for chronic myelogenous leukemia (CML) has been attempted (IV) (127). Roswell Park Memorial Institute in Buffalo, New York reported a series of 15 patients with CML given BCG plus allogeneic cells in addition to maintenance chemotherapy. The cells were thought to be leukemia cells in culture, but later proved to be transformed lymphoblasts. These patients were compared to 28 selected historical controls from the same institution. The survival rate of the immunotherapy group appeared to be improved over the controls, 60% and 20% at five years, respectively ($P < 0.01$). However, this was a small number of patients with chronic disease receiving variable treatment over a long period of time. This combination should be tested in a larger randomized trial. Passive transfer of specifically cytotoxic serum to three patients with chronic lymphocytic leukemia (CLL) caused a drop in leukocyte counts and a reduction in lymph node size, but the response was variable and therapeutic benefit was not obvious (VI) (128).

Overall, the leukemia studies, particularly for AML, indicate that combination treatment with nonspecific and specific immunotherapy has merit and that specific stimulation with tumor cells may be necessary to the treatment. Moreover, it is apparent that immunostimulation works in combination with chemotherapy. That these results were obtained in patients in remission reinforces the concept that immunotherapy is best used when tumor burden is small. While several studies indicate that immunotherapy may help in the treatment of leukemia, the benefit has not been universal, and considerable testing in carefully controlled trials is necessary before its place in the treatment of hematologic malignancies can be determined.

Lymphomas

Burkitt's lymphoma clearly has tumor-associated antigens related to the Epstein-Barr virus (129). However, the results of immunotherapy for this disease have been disappointing (Table 9). A prospective, randomized study in which

Table 8 Other leukemia studies

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
Chronic myelogenous leukemia				
1973	Sokal et al	127	BCG, lymphoid cells	IV
Chronic lymphocytic leukemia				
1968	Lazlo et al	128	Passive transfer of serum	VI
Mixed leukemia studies				
1962	DeCarvalho	124	Xenogeneic antileukemia serum	V
1966	Schwarzenberg et al	119	Adoptive transfer leukocytes	V
1976	Gutterman et al	73	BCG	VI
1976	Bodey et al	116	BCG	VI
1976	Aleksandrowicz & Skotnicki	123	Thymic humoral factor	VI
1977	Hewlett et al	113	BCG, chemotherapy	VII

BCG was given to patients with Burkitt's lymphoma in remission following chemotherapy showed no apparent therapeutic benefit (VII) (130). Relapse rate and duration of remission were the same in 21 immunotherapy patients and 19 control patients. Objective tumor regressions were reported in two patients with Burkitt's lymphoma treated with serum from patients in remission (VI) (131, 132). However, in a controlled study, Fass and colleagues (133) found no effect from the passive transfer of serum from patients in remission (VI), even though the serum had activity against Epstein-Barr virus antigens in vitro. An additional study of a single patient yielded a similar negative result (VI) (134).

Sokal and colleagues (135) reported a study of survival of lymphoma patients given a single BCG vaccination in addition to their conventional therapy (VI). Fifty patients with Stage IA or IIA were randomized to receive BCG vaccination or not. The relapse rate and duration of remission appeared better in the BCG-vaccinated group. While prognostic factors appeared similar in the two groups, there were a number of uncontrolled variables in the study, including various histologic diagnoses and other treatments. Given the number of variables, this study was too small to answer the question posed.

Preliminary results of a large, randomized, controlled trial that tested intradermal BCG added to standard chemotherapy for the treatment of non-Hodgkin's lymphoma have been reported by Jones & Salmon (136). The number of complete remissions achieved was not improved by BCG treatment, but there was a slight increase in the number of partial remissions obtained (not statistically significant). In the maintenance phase, the number of relapses was the same in the BCG group and the no-maintenance drug group. Thus, nonspecific immunotherapy does not appear to be additive in this setting.

In spite of the antigenicity of lymphomatous tumors and their apparent susceptibility to immune attack in transplantation systems, benefit from immunotherapy in clinical trials remains to be demonstrated. It is possible that the addition of active-specific immunostimulation would be helpful.

Lung Carcinoma

Several reports indicate that immunotherapy subsequent to primary treatment for early lung carcinoma is beneficial in reducing recurrences (Table 10). The

Table 9 Lymphoma

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
1967	Burkitt	131	Passive transfer of serum	VI
1967	Ngu	132	Passive transfer of serum	VI
1970	Fass et al	133	Passive transfer of serum	VI
1970	Bluming & Serpick	134	Passive transfer of serum	VI
1974	Sokal et al	135	BCG vaccination, chemotherapy	VI
1976	Magrath & Ziegler	130	BCG	VII
1977	Jones & Salmon	136	Intradermal BCG, chemotherapy	VI

most dramatic of these was the study by McKneally and co-workers (137) of a series of patients with operable lung carcinoma who were randomized to receive either one postoperative dose of intrapleural BCG given via chest tube or no immunotherapy (II). Their study was based partly on the observation that patients who get empyema after operation for lung cancer seem to survive longer (138, 139). All patients with operable lung carcinoma were entered into their study, though oat cell carcinoma was excluded. Results in the group with Stage I disease were striking. At one year follow-up, 17 BCG-treated patients had no deaths and no recurrences. In contrast, 22 patients receiving no additional treatment had nine recurrences and five deaths at one year. This study has been continued, and at a median follow-up of 640 days, 26 of 28 BCG patients (93%) are free of disease compared to 22 of 33 nonimmunotherapy patients (67%) ($P =$ .009) (140). Immunotherapy had very little effect in Stage II or Stage III disease. Thus, for early disease, this approach may be a major improvement in treatment.

Single-dose intradermal BCG following operation for lung carcinoma did not appear to be effective (VI) (141). Sixty BCG-treated patients were compared to a historical group of 60 patients who received surgery alone. There was a slight improvement in survival for the immunotherapy group, but it was not statistically significant. For successful immunotherapy in experimental tumor models, it is critical to get the immunostimulant into the regional lymphatics draining the tumor (18). This finding appears to be confirmed by the benefit from intrapleural BCG in McKneally's study compared to the failure of intradermal BCG in the same setting. However, McKneally was administering a larger dose of BCG than was given intradermally.

Table 10 Lung carcinoma

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
Adjuvant immunotherapy following primary therapy				
1974	Edwards & Whitwell	141	Intradermal BCG, surgery	VI
1975	Amery et al	40	Levamisole, surgery	VI
1976	McKneally et al	137	Intrapleural BCG, surgery	II
1976	Pines	143	Intradermal BCG, radiation	II
1977	Stewart et al	145	Tumor antigen, chemotherapy, and surgery	II
1977	Takita et al	144	Neuraminidase-treated cells, ConA, CFA, and surgery	II
Treatment for disseminated disease				
1973	Alth et al	147	Tumor extract	VI
1975	Israel & Edelstein	33	*C. parvum*, chemotherapy	II
1976	Kimura et al	148	Strep OK-432, chemotherapy	VI
1977	Yamamura	155	Oil-attached cell-wall skeleton BCG	VI
1977	Dimitrov et al	146	*C. parvum*, chemotherapy	VI

A multicenter, prospective, randomized trial using levamisole as adjuvant therapy for resectable lung carcinoma has been reported (VI) (40). Oat cell carcinoma was excluded. The immunotherapy patients were given levamisole three days preoperatively, and 150 mg orally for three days every two weeks thereafter; the others were given a placebo. The published results of a one year follow-up showed a trend toward a decreased number of recurrences and increased survival time in favor of the levamisole group: 10 recurrences in 51 levamisole patients (20%) compared to 20 of 60 patients receiving a placebo (33%), but this difference is not statistically significant. It was interesting, however, that the recurrence rate in the levamisole group remained the same for both large and small primary tumors, while it increased markedly for larger tumors in the placebo group. In a more recent analysis of the data, Amery (142) noted that important differences were observed if the test and control groups were divided into patients weighing more or less than 70 kg. In the lighter patient group there were fewer recurrences or deaths, and this difference was highly significant when compared to patients receiving the placebo. In the heavier group, these differences did not exist. This finding suggests that immunopotentiation with levamisole is dose-related, and, in appropriate doses, it may be an effective adjuvant therapy. This must be demonstrated in an appropriately stratified trial however.

A group of patients with inoperable lung carcinoma confined to the chest was treated by irradiation followed by adjuvant immunotherapy (II) (143). The patients were randomized to receive postradiation BCG or no additional treatment. Minimum follow-up was two years. During the first year, 12 of 23 patients (52%) receiving no additional treatment died compared to five of 25 patients (20%) receiving BCG. The BCG group had a significantly better one year survival rate ($P < .02$), although after the first year of follow-up, survival rates in the two groups were similar. Patients receiving only palliative radiation were not entered in this group, but it was observed that they did not respond to BCG treatment. Those receiving a full course of radiation were apparently benefited initially by adjuvant immunotherapy.

A clinical trial of active-specific immunotherapy as an adjuvant to surgery and radiation for locally advanced lung carcinoma has been reported (VI) (44). All gross disease was removed surgically in 11 patients. If tumor remained in the bronchial stump or on the chest wall, radiation was given. By random selection, five patients received a vaccine of autologous tumor cells treated with concanavalin A suspended in Freund's adjuvant. The other six patients received no additional treatment until recurrence developed. Chemotherapy was used in both groups if recurrence appeared. Three of the five immunotherapy patients were alive with a median survival time for the group of 15 months at the time of the original study. All six patients in the control group died, with a median survival time of five months. An update of this study has been reported with 15 patients in each group (II) (144). In the immunotherapy group, seven patients are alive, the longest at 51 months, with a median survival time for the group of

34.8 months. In the control group, three patients are alive with the longest survivor at 26 months and a median survival for the group of 12.1 months. This difference in survival is statistically significant ($P < .05$).

Stewart and colleagues (145) have reported preliminary results from another study of patients with lung cancer who were treated with tumor antigen in complete Freund's adjuvant and chemotherapy following surgery (II). This was a prospective, randomized trial that compared adjuvant chemo-immunotherapy to chemotherapy alone. The data indicated an increase in survival time for the combined adjuvant therapy patients that was significantly better than for patients given chemotherapy alone ($P = .001$). Both the Takita and Stewart studies suggest benefit from adjuvant chemo-immunotherapy. However, relatively small numbers of patients were entered into the trials over a long period, which suggests some selection of patients and this may have influenced their findings.

Adjuvant immunotherapy for lung carcinoma is promising, especially for patients in the earlier stages of the disease. All studies of adjuvant treatment emphasize the need for a significant reduction of tumor burden before immunotherapy can be effective. Additional studies are warranted.

Immunotherapy has been used in the treatment of disseminated lung carcinoma. Israel & Edelstein (33) compared subcutaneous *C. parvum* combined with five-drug chemotherapy to chemotherapy alone in a prospective randomized trial (II). The median survival time for 75 patients receiving chemotherapy was only four months; in the chemo-immunotherapy group, 68 patients showed a median survival time of nine months, a significant improvement ($P < .001$). A similar increase in survival time was reported when combined chemo-immunotherapy was used for disseminated oat cell carcinoma. The immunotherapy group had a mean survival time of 9.1 months compared to five months for the control group ($P < .02$) (33). The benefit from *C. parvum* in this setting has not yet been confirmed by other investigators. Preliminary results from a prospective, randomized trial reported by Dimitrov and colleagues (146) showed no improvement in survival time (over chemotherapy alone) when patients received a combination of adriamycin and subcutaneous *C. parvum* for advanced lung carcinoma (VI). However, these are early results from a trial employing only single-agent chemotherapy.

A small, uncontrolled study using tumor cell extract as the sole treatment for disseminated lung carcinoma was reported by Alth and colleagues (147), but therapeutic effect appeared to be minimal (VI). A streptococcal agent, OK-432, was combined with chemotherapy for disseminated lung carcinoma by Kimura and co-workers (148) (VI). A slight increase in the median survival time was achieved in the treated group over a chemotherapy control group, but the data did not establish this as clearly beneficial. A group of patients with advanced lung carcinoma were given the oil-attached cell-wall skeleton of BCG in addition to conventional therapy (VI) (155). The 50% survival interval for patients with Stage III disease was 18 months, as opposed to seven months for patients treated by conventional therapy alone. Although this result was statistically significant, the latter group was comprised of historical controls, and it was not clear whether the conventional therapy in the treatment and the control group was comparable.

Despite the shortcomings of some clinical trials, it would appear that immunotherapy can be beneficial to patients with lung carcinoma. Immunotherapy must be combined with other agents that significantly reduce tumor burden. Additional randomized, controlled studies are needed, both for adjuvant treatment and for combination therapy for disseminated disease, if immunotherapy is to become an established treatment modality for lung carcinoma.

Malignant Melanoma

LOCAL IMMUNOTHERAPY Results obtained in the treatment of malignant melanoma have been the major stimulus to the use of immunotherapy for other solid tumors. In immunocompetent patients, cutaneous metastases of malignant melanoma will consistently undergo complete regression following intralesional injection with BCG. This observation, originally reported by Morton and his colleagues (12), has been confirmed by numerous investigators (I-A) (156–160, 150–152, 63). Serial biopsies of injected lesions show an intense lymphocytic and mononuclear cell infiltrate and destruction of melanoma cells. Moreover, in 17% of patients receiving intralesional BCG, there is a concomitant regression of uninjected nodules (12, 156, 159, 151, 63). Approximately 20% of patients responding to this treatment remain free of disease over a long period; the longest is still living at ten years (153). However, most patients develop recurrent lesions and eventually die of metastatic melanoma. It is not known whether the intralesional treatment prolongs survival time in these patients. During the treatment of metastatic lesions with intralesional BCG, new lesions will develop in a significant number of patients. These may be treated as well. The treatment is ineffective in patients who remain anergic to both DNCB and PPD skin testing. If the tumor burden is large or visceral metastases are evident, local responses in the skin are less frequent. Clinically detectable visceral metastases very rarely regress concomitantly with skin lesions. However, for metastatic disease limited to the skin or for small subcutaneous nodules, the response to intralesional injection is consistent and may be considered a useful, alternative therapy. Occasional responses in distant nodal or visceral disease are seen, but not often enough to be clinically useful.

Other types of immunotherapy have been used successfully to control accessible metastases of malignant melanoma. The induction of local, delayed hypersensitivity responses to DNCB can be effective in eliminating cutaneous disease (I-B) (161, 162). Malek-Mansour (162) reported that skin metastases in five of seven patients with malignant melanoma completely regressed after DNCB treatment of the lesions (I-B). Two of these remained free of disease from one to three years.

Intralesional injection with smallpox vaccine (vaccinia virus) will destroy intradermal deposits of malignant melanoma (163–165). Hunter-Craig and his colleagues (165) reported that skin metastases regressed completely in six of ten patients following virus injection (I-B). Most of the patients had been followed less than ten months, but two were free of disease at 11 and 22 months after injection. Responders and nonresponders could not be separated on the basis of clinical findings. Responses were limited to the injected lesions, and subcutaneous or nodal disease did not respond. Side effects were those of smallpox vaccination. As another approach, mumps virus has been added to intralesional BCG injection, but it is not clear from the published data whether it augmented the response (VI) (166). Injection of mumps virus alone will cause regression of lesions (I-B) (152).

Intralesional injections of autochthonous lymphocytes stimulated with PHA-induced partial regressions of cutaneous melanoma, but did not eliminate the lesions (I-B) (68). Smith and his colleagues (63) reported that intralesional injection of transfer factor from persons sensitized to the patient's tumor caused regression of skin metastases of malignant melanoma (V). He noted one complete regression lasting more than two years. *C. parvum* has been injected into cutaneous metastases of malignant melanoma (V) (167). Six of 14 patients injected showed regression of lesions (three complete, three partial). There was no regression of noninjected lesions. Although some success has been achieved with all of these treatments, BCG has been the agent most tested and used for local immunotherapy of malignant melanoma (See Table 11).

ADJUVANT IMMUNOTHERAPY Adjuvant immunotherapy may be effective for reducing the high risk of recurrence in certain patients following surgical treatment for malignant melanoma. Rationale for this form of therapy is suggested by several observations: (*a*) the concomitant regression of uninjected nodules in patients receiving intralesional BCG; (*b*) the increase in antimelanoma antibody after local immunotherapy; and (*c*) the prolonged survival time of patients treated with BCG for skin metastases who, based on the natural history of metastatic malignant melanoma, have a high likelihood of subclinical, visceral metastases as well.

A controlled trial using BCG as adjuvant therapy following surgery for malignant melanoma metastatic to the regional nodes was reported by Eilber and colleagues from UCLA (III) (45). Patients with Stage II disease treated by operation alone are known to have a 70% recurrence rate within two years, and 80% recur within five years. In Eilber's study, 126 patients entered after radical resection of regional nodes and histologic confirmation of stage of disease. Of these, 84 patients received immunotherapy consisting of weekly intradermal BCG (3×10^8 organisms by Tine technique) for 12–52 weeks, and then every two weeks thereafter for two years; 42 were treated by operation alone. The results were evaluated by life table analysis, and showed a marked improvement in the probability of patients remaining free of disease at two years postoperation. In the BCG-treated group, 64% of the patients were free of disease vs 36% in the group treated by operation alone ($P < .05$). In those who recurred, median time to recurrence in the BCG group was ten months, whereas for those treated by operation alone, the median time was four months. Whether the number of survivors will ultimately be greater with BCG therapy must await longer follow-up.

Immunotherapy was more effective in patients with a smaller tumor burden. Of patients receiving BCG for one lymph node (positive only on microscopic exam), 90% were disease-free at two years, while only 62% of those with a clinically palpable node remained free of disease on immunotherapy. In the surgery-only group, recurrence rate was the same for patients with microscopic or palpable nodal metastases, 36% and 32%, respectively. This study was not strictly randomized, but the immunotherapy and control groups were very similar, except for a larger number of patients over age 40 in the control group. A completely randomized, prospective trial using adjuvant immunotherapy for Stage II malignant melanoma is now in progress (167a).

Another controlled trial of adjuvant immunotherapy following operation for

Table 11 Malignant melanoma

Year	Investigator(s)	Ref.	Local immunotherapy for cutaneous or accessible disease	Type of study
1970	Morton et al	12	Intralesional BCG	I-A
1972	Nathanson et al	156	Intralesional BCG	I-A
1972	Seigler et al	157	Intralesional BCG	I-A
1973	Baker & Taub	158	Intralesional BCG	VI
1973	Pinsky et al	159	Intralesional BCG	I-A
1973	Smith et al	63	Intralesional BCG	V
1974	Grant et al	160	Intralesional BCG	I-A
1974	Lieberman et al	149	Intralesional BCG	V
1974	Morton et al	153	Intralesional BCG	I-A
1975	Israel et al	150	Intralesional BCG	I-A
1976	Karakousis et al	154	Intralesional BCG	VI
1976	Mastrangelo et al	151	Intralesional BCG	V
1976	Gerner & Moore	152	Intralesional BCG & PPD, mumps, Strep	I-B
1972	Klein & Holtermann	161	Topical DNCB	V
1973	Malek-Mansour	162	Topical & intralesional DNCB	I-B
1964	Burdick & Hawk	163	Intralesional vaccinia virus	VI
1966	Milton & Lane Brown	164	Intralesional vaccinia virus	V
1970	Hunter-Craig et al	165	Intralesional vaccinia virus	I-B
1973	Minton	166	Intralesional mumps virus, BCG	VI
1972	Cheema & Hersh	68	Intralesional activated lymphocytes	I-B
1973	Smith et al	63	Intralesional transfer factor	V
1975	Cunningham-Rundles et al	167	Intralesional *C. parvum*	V

malignant melanoma metastatic to the regional nodes has been reported from M. D. Anderson Hospital and Tumor Institute (IV) (73). An important difference, however, between this and the first study is that the UCLA group removed the lymph nodes prophylactically for staging, while at M. D. Anderson lymph node dissection was not done until the nodes were clinically positive. This means in this study immunotherapy was introduced later in the course of the disease in some patients. The treatment group was randomized to receive different doses of BCG. Results were compared to historical controls treated at M. D. Anderson by operation alone. This study showed that the number of recurrences was the same in the treatment and control groups, but those patients receiving high-dose BCG (6×10^8 viable organisms) had a longer disease-free interval and a longer survival ($P = .002$). Low-dose BCG (6×10^7 viable organisms) had essentially no effect on the course of the disease, which suggests a dose relation to immunostimulation. This dose dependence has been observed in experimental animal studies (18). A similar adjuvant study from Sloan-Kettering (New York) in which immunotherapy was introduced after resection of clinically involved regional nodes showed no difference in recurrence rate between treatment and control groups and no increase in survival rate in the immunotherapy group at two years follow-up (VII) (168). These investigators used a lower dose of BCG ($4\text{–}6 \times 10^7$ organisms) however. Each of these studies suggest that the dose of BCG may be critical to the success of the treatment.

In the adjuvant study from M. D. Anderson, results showed no benefit from BCG therapy for patients with malignant melanoma of the head and neck (73). These investigators thought that their lack of success might be due to failure of the BCG to reach appropriate regional lymphatics of the head and neck. A report from UCLA indicated a trend in favor of BCG therapy for reducing the number of recurrences following surgery for head and neck melanoma, but this difference was not statistically significant (IV) (169). However, there was a significant increase in the disease-free interval (from four to 15 months) for immunotherapy patients over those treated by surgery alone ($P < .001$), and length of survival also was significantly increased ($P < .05$).

The potential usefulness of adjuvant immunotherapy for malignant melanoma makes it imperative that its efficacy be established with certainty. Evidence must come from well-conducted, prospective, randomized trials. The other adjuvant studies listed in Table 12 are either preliminary work, uncontrolled studies, or they compare results to historical controls obtained from other published series.

COMBINATION CHEMO-IMMUNOTHERAPY For the treatment of disseminated malignant melanoma, nonspecific immunostimulation has been added to chemotherapy (Table 13). Gutterman and colleagues (179) reported a controlled trial using combined dimethyl-triazeno imidazole carboxamide (DTIC) and intradermal BCG vs DTIC alone (historical controls were treated by the same group of investigators at M. D. Anderson) (III). The overall response rate was 27%

Table 12 Malignant melanoma

Year	Investigator(s)	Ref.	Adjuvant immunotherapy following surgery	Type of study
1972	Bluming et al	170	Intradermal BCG	VI
1972	Ikonopisov	171	Tumor cells, intralesional BCG	VI
1973	Gutterman et al	172	Intradermal BCG	VI
1974	Morton et al	153	Intradermal BCG	VI
1975	Serrou et al	173	Tumor cells, intradermal BCG	VI
1975	MacGregor et al	174	Oral BCG	VI
1975	Thompson	175	Intradermal BCG	VI
1976	Karakousis et al	154	Intradermal BCG	VI
1976	Jewell et al	52	Exchange of tumors and leukocytes	VI
1976	Gutterman et al	73	Intradermal BCG	IV
1976	Eilber et al	45	Intradermal BCG	III
1976	Eilber et al	169	Intradermal BCG, head and neck	IV
1976	Gerner & Moore	152	Intradermal BCG, PPD, tumor cells	VI
1977	Cunningham et al	176	Intradermal BCG, DTIC	VI
1977	Pinsky et al	168	Intradermal BCG	VII
1977	Beretta	177	Intradermal BCG, DTIC	VI
1977	Spitler et al	178	Levamisole	VI

(6% complete responses) for combination therapy compared to 14% for DTIC alone. In patients without visceral metastases, the response rate was 55% with combination therapy vs 18% for drug only ($P = .025$). Overall, the duration of remission and length of survival was increased with chemo-immunotherapy over chemotherapy alone ($P = .05$). In another study a good response rate was obtained with DTIC and vincristine plus intradermal BCG and radiated allogeneic melanoma cells, but the data reported were preliminary and therapeutic benefit could not be evaluated (VI) (180).

Table 13 Malignant melanoma

Year	Investigator(s)	Ref.	Combination chemo-immunotherapy for disseminated disease	Type of study
1974	Gutterman et al	179	DTIC, intradermal BCG	III
1975	Currie & McElwain	180	DTIC, vincristine, BCG	VI
1975	Israel & Edelstein	33	Five-drug therapy, *C. parvum*	VI
1976	Gutterman et al	182	DTIC, MeCCNU, BCG	IV
1976	Newlands et al	181	DTIC, ICRF 159, BCG	VII
1977	Costanzi	183	Triple-drug chemotherapy, BCG	VII
1977	Mastrangelo et al	184	MeCCNU, vincristine, BCG	VII
1977	Presant et al	185	Cytoxan, DTIC, *C. parvum*	VII

A randomized, controlled trial comparing DTIC, 1,2-d:(3,5 dioxopiperazin-1-yl)propane (ICRF 159) plus intradermal BCG and radiated allogeneic melanoma cells has been reported (VII) (181). This study included patients with disseminated disease and those with only regional lymph node metastases. This latter group would be treated surgically in many centers. For disseminated disease, there was no statistically significant difference between chemo-immunotherapy and drugs alone. However, the number of patients in the trial was relatively small for detection of subtle differences. Investigators from M. D. Anderson Hospital and Tumor Institute reported that the addition of (1-(2-chloroethyl)-3-4-methylcyclohexyl)-1 nitrosourea (MeCCNU) to the DTIC-BCG combination gave results equal to the previous trial with DTIC-BCG, but no better (IV) (182).

A prospective, randomized trial comparing 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU), hydroxyurea, and DTIC with or without BCG to the DTIC-BCG combination has been reported (VII) (183). The response rate was better with the triple-drug therapy than with the DTIC-BCG combination; 28% vs 20%, respectively. Otherwise, survival rate was the same. The addition of BCG did not improve the results of triple-drug therapy. In a similar trial, the addition of BCG did not improve results obtained from MeCCNU and vincristine (VII) (184).

The addition of *C. parvum* to combination chemotherapy for patients with disseminated malignant melanoma has been reported to improve survival time (VI) (33). These results were from a nonrandomized trial, however. A recent prospective, randomized trial did not show benefit from intravenous *C. parvum* in this situation (185). Overall, the treatment for disseminated malignant melanoma is unsatisfactory. The role of immunotherapy given in combination with chemotherapy is uncertain, and the chemotherapy itself is relatively ineffective for this disease.

OTHER MELANOMA STUDIES Passive immunotherapy has been used to treat malignant melanoma with some success (Table 14). Sumner & Foraker (49) transfused whole blood from a patient with malignant melanoma who had undergone complete spontaneous regression to two other melanoma patients (VI). One patient developed complete regression of his disease following transfusion. The adoptive transfer of specifically sensitized, immunocompetent cells has been tested extensively in clinical trials (186–188, 50, 52). The procedure involves matching patients with similar tumors, and then cross transplanting the tumor cells. Leukocyte transfusions are then exchanged between the partners. The overall response rate in patients with disseminated melanoma is about 20%. Nadler & Moore (50) reported two long-lasting complete regressions of melanoma in 86 patients (V). Krementz (51) reported responses in nine of 56 melanoma patients; three responses were complete for one year to over four years (V). Transfusion incompatibilities have been responsible for some patient deaths with this treatment modality.

Transfer factor has been tested for disseminated malignant melanoma in an effort to avoid some of the problems associated with exchange of intact cells. A recent report by Jewell and associates (52) indicated complete responses in three of nine patients receiving transfer factor from sensitized donors (V). In an earlier report, Spitler (189) reported one complete regression in four melanoma patients treated with transfer factor (VI). In the other studies with transfer factor, there was either no benefit or the therapeutic benefit could not be evaluated from the data given (190–192, 51). In another study (60), xenogeneic immune RNA was used for malignant melanoma, but was of limited benefit (VI). Fifteen patients with advanced melanoma showed no regressions in response to the treatment. Metastatic disease in some of the patients appeared to stabilize for a period of time but eventually progressed.

Active-specific immunotherapy consisting of tumor cell vaccines has not clearly shown benefit in the treatment of malignant melanoma (187, 193, 194). However, specific immunotherapy should be helpful since melanomas appear to have distinct, tumor-associated antigens, and immune responses to these antigens have been clearly demonstrated (7). Clinical trials of BCG combined with allogeneic tumor cell vaccine as an adjuvant to surgical resection of regional disease are in progress. Preliminary results show that patients receiving tumor cell vaccine increase their antimelanoma antibody titer more than those receiving BCG only or no additional treatment (K. Irie, personal communication,

Table 14 Malignant melanoma

Year	Investigator(s)	Ref.	Immunotherapy for established disease	Type of study
1960	Sumner & Foraker	49	Transfusion whole blood	VI
1969	Nadler & Moore	50	Exchange of tumors and leukocytes	V
1971	Curtis	186	Exchange of tumors and leukocytes	VI
1971	Krementz et al	187	Exchange of tumors and leukocytes	V
1973	Goodwin et al	188	Exchange of tumors and leukocytes	VI
1974	Krementz et al	51	Exchange of tumors and leukocytes	V
1976	Jewell et al	52	Exchange of tumors and leukocytes	V
1971	Krementz et al	187	Tumor cell vaccine	VI
1973	Seigler et al	193	Neuraminidase treated cells, BCG	VI
1974	Levy et al	194	Neuraminidase treated cells, BCG	VI
1971	Brandes et al	190	Transfer factor	VI
1972	Spitler et al	189	Transfer factor	VI
1973	Morse et al	191	Transfer factor	VI
1974	Krementz et al	51	Transfer factor	VI
1974	Price et al	192	Transfer factor	VI
1976	Jewell et al	52	Transfer factor	V
1977	Ramming et al	60	Xenogeneic immune RNA	VI
1972	Ghose et al	195	Chlorambucil-carrying antibody	VI
1973	Falk et al	196	Oral BCG	VI
1975	MacGregor et al	174	Oral BCG	VI

1977). The increase in antimelanoma antibody titer correlates very well with one year survival. Irrespective of the role of antibodies in controlling tumor, their presence indicates that specific responses may be important to the success of immunotherapy. It is possible that combinations of specific immunotherapy, nonspecific immunostimulation and potent drug therapy could improve the treatment of disseminated disease. At least, in certain situations, immunotherapy for malignant melanoma offers promise but much improvement and investigation are still needed.

Sarcomas

Osteosarcomas are aggressive tumors that cause death from pulmonary metastases, despite adequate surgical control of the primary tumor (Table 15). Sarcomas carry well-defined, tumor-associated antigens, and stimulate immune responses in their autochthonous hosts (197). An early report by Marcove and associates (198) suggested that active immunization with lysed sarcoma cells was helpful in prolonging the disease-free interval following surgery for osteosarcoma (VI). These were preliminary results, however, and the study was uncontrolled. In another adjuvant study reported by Marsh and co-workers (199), tumor removed at the time of amputation was transplanted to another patient with osteosarcoma (VI). The tumor-implant recipient's lymphocytes were then transfused to the tumor-implant donor. Preliminary data, again, were suggestive of benefit, but a follow-up report on 32 patients treated in this fashion

Table 15 Sarcomas

Year	Investigator(s)	Ref.	Tumor Type	Immunotherapy	Type of study
1971	Marcove et al	198	Osteosarcoma	Tumor cell vaccine	VI
1972	Marsh et al	199	Osteosarcoma	Exchange of tumor and lymphocytes	VI
1975	Neff & Enneking	200	Osteosarcoma	Exchange of tumor and lymphocytes	VII
1975	Eilber et al	201	Osteosarcoma	BCG, tumor cell vaccine	VII
1975	Levin et al	55	Osteosarcoma	Transfer factor	VI
1976	LoBuglio & Neidhart	56	Osteosarcoma	Transfer factor	VI
1976	Green et al	208	Osteosarcoma	Viral oncolysates	VI
1977	Ritts et al	203	Osteosarcoma	Transfer factor	VI
1970	Morton et al	205	Mixed	BCG, tumor cell vaccine	VI
1971	Morton et al	206	Mixed	BCG, tumor cell vaccine	VI
1973	LoBuglio et al	204	Alveolar soft part	Transfer factor	VI
1974	Gunnarson et al	209	Mixed	Chemotherapy, *C. parvum*	VI
1976	Townsend et al	47	Soft tissue	BCG, tumor cell vaccine	VI
1977	Sinkovics	207	Mixed	BCG, viral oncolysates, chemotherapy	VI

by Neff & Enneking (200) indicated that the number of recurrences and the length of disease-free interval were the same as in a group of historical controls treated by surgery alone (VII). A controlled trial of adjuvant immunotherapy consisting of BCG and allogeneic sarcoma cells following surgery for osteosarcoma was reported by Eilber and colleagues (201). In this study, adjuvant immunotherapy did not decrease recurrence or increase survival time, even in Stage I disease. As a group, these papers suggest that active immunization and/or BCG as the only adjuvant to surgery do not help the osteosarcoma patients.

A report by Levin and colleagues (55) dealt with the use of transfer factor (TF) for osteosarcoma (VI). TF, obtained from the lymphocytes of relatives and household contacts, showed in vitro activity against the osteosarcoma patient's tumor. Six patients treated with TF after amputation for osteosarcoma (no metastases were noted at the time of surgery), were alive and free of disease 9–26 months after operation (median, 14 months). LoBuglio & Neidhart (56) reported an uncontrolled trial with five patients receiving TF as an adjunct to surgical resection for osteosarcoma (VI). Two patients recurred at 4 and 12 months; three were disease-free at 7, 8, and 12 months. For comparison, Eilber (201) reported a median recurrence time after operation of 3.1 months. This longer disease-free interval with TF is promising, but adjuvant studies cannot be evaluated in the absence of controls. A report from the Mayo Clinic suggested that there has been an improvement in survival from osteosarcoma unrelated to any therapeutic modality. This finding underscores the need for concomitant controls in testing any new agents (202). An interim report from a trial comparing TF with methotrexate, adriamycin, and vincristine given as adjuvant therapy following operation for osteosarcoma has appeared (VI) (203). No clear difference existed between the two groups at 20 months follow-up.

Osteosarcoma, even Stage I, has a dismal prognosis. These tumors grow so rapidly that the tumor burden from occult metastases may be too great even in the subclinical state for host resistance to have any effect. Potent cytoreductive therapy to lower tumor burden is essential with this type of cancer. Drugs with high activity against osteosarcoma are available. and, when given as adjuvant to surgical procedures, appear to reduce recurrences and increase survival time. Immunotherapy might be useful in maintaining remission in conjunction with the drugs or beyond the period of adjuvant chemotherapy. LoBuglio & Neidhart (56) mentioned two patients with metastatic disease whose tumors regressed on adriamycin and DTIC, and who remained in remission for 11 and 28 months on maintenance therapy with TF.

Immunotherapy has been used in the treatment of various soft-tissue sarcomas (VI) (205, 206, 47). Following surgical removal of the tumor, the patients in the latest study received either BCG plus allogeneic sarcoma cells or no additional treatment. Follow-up was two years. For Stages I and II disease, there were fewer recurrences in patients receiving immunotherapy. In this group, 61% of the patients remained free of disease vs 33% of those treated by operation alone. In those who recurred, the median disease-free interval was 12 months for the immunotherapy group, and six months for those treated by operation alone. These figures were not statistically significant, however. Immunotherapy did not benefit those with more advanced disease. Potent cytoreductive therapy is equally important for soft-tissue sarcomas. Adjuvant chemotherapy has come into use for soft-tissue sarcomas as well, and clinical trials combining immunotherapy with chemotherapy are in progress.

Immunotherapy has been used for the treatment of disseminated sarcomas. LoBuglio and associates (204) treated a patient with metastatic alveolar soft-part sarcoma with TF obtained from his twin brother (VI). The tumor did not progress during the period of treatment, although other factors also may have contributed to the remission. Israel & Edelstein (33) combined subcutaneous *C. parvum* with five-drug chemotherapy for treatment of disseminated, nonlymphomatous sarcomas in a randomized trial (II). Median survival time was longer by several months in the *C. parvum* chemo-immunotherapy group than in those receiving drugs alone ($P. < .03$). There have been no other studies to confirm this result. Sinkovics (207) recently noted (in a preliminary report) that the addition of viral oncolysate to BCG and chemotherapy for the treatment of disseminated sarcomas improved the response rate (VI). The use of viral oncolysates alone had no effect in 12 patients with disseminated osteosarcoma (VI) (208).

The marked antigenicity of sarcomas, combined with the ready demonstration of immune responses to tumor in sarcoma patients, suggests that immunotherapy should be helpful. However, such a role has yet to be defined.

Skin Carcinoma

Klein (13, 210) reported that skin cancers could be eradicated with local immunotherapy, and provided part of the stimulus for the current interest in immunotherapy for malignant disease (Table 16). He described the induction of local DCH responses by various agents to eliminate a high percentage of treated lesions. This included patients with multiple, superficial, basal cell carcinomas, squamous cell carcinomas in situ, premalignant keratoses, and leukoplakia. Elimination of these lesions was accomplished without systemic toxicity or damage to normal skin.

Table 16 Carcinoma of the Skin

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
1968	Klein	13	Topical TEIB or DNCB	I-A
1969	Klein	210	Topical TEIB or DNCB	I-A
1970	Williams & Klein	212	Topical 5-FU	I-A
1971	Stjernsward & Levin	67	Topical DNCB	I-A
1973	Levis et al	211	Topical DNCB	I-A
1976	Raaf et al	213	Topical DNCB and 5-FU	VI
1976	Klein et al	64	Various topical and intralesional agents	I-A

For the treatment of basal cell carcinomas, Klein sensitized five patients to TEIB (2,3,5-triethylene imino-1,4-benzoquinone). TEIB was then applied topically to multiple, superficial basal cell carcinomas. He reported a 95% success rate for these lesions (I-A) (13, 210). This claim was supported by a description of a dramatic result in a single patient with 140 lesions. The individual response rate in other patients, or the number of individual lesions treated, was not described, so expectations for individual patients were not clear. Levis and colleagues (211) reported a much lower success rate for treatment of basal cell carcinomas (I-A). They used DNCB topically to treat five patients with 113 lesions, and found that only 32% of the lesions completely regressed. An additional 29% showed partial regression. Another patient, not included in their data, had uncountable lesions but obtained a good response to DNCB treatment. Individual response rates for complete regression varied from 22 to 70% of the lesions. They were treating both superficial and nodular basal cell disease, and the response was equal for both, but nodular lesions measuring greater than 5 mm did not respond. Stjernsward & Levin (67) reported that at least one basal cell carcinoma regressed completely in eight of nine patients treated with topical DNCB (I-A). They observed, however, that not all lesions responded, and they did not give the actual number of lesions treated or the overall response rate. For comparison, Williams & Klein (212) have reported that topical 5-flourouracil (5-FU) produced resolution of 56% of nodular basal cell carcinomas (I-A). This result is inferior to surgical excision. Thus, for patients with multiple lesions that are not suitable for conventional therapy, local immunotherapy may be tried. However, for single or large lesions, excision is the treatment of choice.

Klein (13, 210) has used topical immunotherapy for treating superficial squamous cell carcinomas in patients with xeroderma pigmentosum and arsenical dermatitis (I-A). Again, he reports a high success rate (90%) in eradicating these lesions. Seven patients with xeroderma pigmentosum and six patients with arsenical dermatitis were treated with either topical TEIB or DNCB. No analysis of the data or breakdown for individual patients was given. He observed that this high success rate was limited to superficial lesions. Deeper or recurrent lesions responded less well, although benefit could be obtained by the addition of topical 5-FU. Topical 5-FU alone reportedly eliminated 80% of superficial squamous cell carcinomas (I-A) (212). However, this treatment also had a lower success rate for deeper lesions. In xeroderma pigmentosum, the appearance of new lesions seemed to decrease with topical immunotherapy. Raaf and colleagues (213) reported a single case in which topical DNCB was combined with topical 5-FU for treating a large, squamous cell carcinoma resulting from arsenic exposure (VI). This lesion was refractory to DNCB alone, but regressed completely with both agents. In this patient, 60 smaller lesions regressed completely with topical DNCB alone. Klein, too, advocates the combined treatment.

According to Klein, certain types of premalignant skin lesions can be eliminated by topical immunotherapy. This treatment caused 95% of multiple solar keratoses in 18 patients to completely resolve (13, 210). Arsenical keratoses were refractory to treatment, however, and did not respond. In 14 patients with leukoplakia, good initial responses could be obtained, but the lesions recurred in four to six weeks. Repeated treatment reduced the incidence of recurrence, but Klein gave no data on individual patients.

Thus, for malignant and premalignant lesions of the skin, high success rates have been reported for topical immunotherapy. However, the published reports have emphasized the dramatic successes, and given little data on which to base expectations for individual patients. The study by Levis and colleagues is an exception. Nevertheless, topical immunotherapy appears to be a useful tool for treating superficial skin malignancies and is well worth a trial when conventional treatment is not suitable.

Miscellaneous Advanced Cancers

In clinical trials where immunotherapeutic agents have been tested against a variety of advanced cancers, the results generally have not been striking. However, many of these studies (Table 17) represent interesting techniques or concepts, and some good responses have been obtained. When nonspecific immunostimulation with BCG was added to chemotherapy, Villasor (214) re-

Table 17 Miscellaneous advanced cancers

Year	Investigator(s)	Ref.	Immunotherapy	Type of study
Nonspecific immunostimulation				
1965	Villasor	214	BCG, chemotherapy	VI
1974	Asada	215	Mumps virus	VI
1975	Israel & Edelstein	216	*C. parvum*	V
Active specific immunostimulation				
1964	Aswaq et al	217	Autologous tumor cell vaccine, complete Freund's adjuvant (CFA)	VI
1967	Czajkowski et al	219	Autologous tumor coupled with rabbit γ-globulin	VI
1969	Cunningham et al	220	Autologous tumor coupled with rabbit γ-globulin	VI
1970	Hughes et al	218	Tumor extract, CFA	VI
1972	Humphrey et al	221	Tumor homogenate	V
1977	Wallack et al	222	Vaccinia tumor cell oncolysate	V
Passive transfer immunotherapy				
1960	Finney et al	223	IM and intralesional injections plasma fractions	VI
1963	Woodruff & Nolan	224	Adoptive transfer spleen cells	VI
1967	Andrews et al	225	Exchange of tumors and TD lymphocytes	VI
1972	Yonemoto & Terasaki	226	Thoracic duct lymphocytes	V
1973	Symes & Riddell	227	Intra-arterial pig lymph node cells	V
1976	Vetto et al	57	Transfer factor	V
1976	Frenster	228	PHA-activated autologous lymphocytes	VI

ported increased responses and survival rates over chemotherapy alone for advanced disease (VI). However, his data were not statistically significant. Asada (215) claimed very good results for advanced cancer using mumps virus inoculation (VI). The criteria for good response and the role of the mumps virus, apart from other treatments the patient received, were unclear, however.

Israel & Edelstein (216) treated advanced cancers that were refractory to other treatments with intravenous *C. parvum* (V). They obtained nine partial responses of 50% or greater in 33 patients. Length of response could not be evaluated because chemotherapy then was resumed in these patients. However, response was better than expected from nonspecific immunostimulation in the presence of established tumor.

Another group of studies dealt with active-specific immunostimulation for the treatment of advanced cancer. Aswaq and colleagues (217) administered autologous tumor cells in complete Freund's adjuvant, but obtained no objective responses (VI). Hughes and colleages (218) used a similar procedure and reported a 10% response rate in 21 patients with advanced disease (VI). Czajkowski and colleages (219) coupled autologous tumor cells to rabbit γ-globulin in an effort to increase immunogenicity of the cells. They reported two objective responses in 14 patients after administration of this vaccine (VI). With the same treatment, Cunningham and co-workers (220) obtained only one response in 35 patients (VI). Humphrey and colleagues (221 reported a 23% response rate in 96 patients following administration of tumor homogenate (V). Data regarding the extent or length of these responses were not given. Wallack and associates (222) used an autochthonous tumor cell vaccine lysed by vaccinia virus in 13 patients with various advanced cancers (V). They noted two short-term responses. In general, active-specific immunostimulation as the sole therapy for advanced malignant disease is of limited value.

The passive transfer of immunotherapeutic agents has been used to treat advanced disease. An early study by Finney and colleagues (223) tested intramuscular and intralesional injections of autologous plasma fractions (VI), but these were of limited benefit. The adoptive transfer of spleen cells from unrelated donors to patients with advanced cancer elicited no objective responses, as reported by Woodruff & Nolan (VI) (224). Andrews and colleagues (225) transferred tumors between patients with similar disease and then exchanged thoracic duct lymphocytes obtained from the partner (VI). They obtained no response in four patients. Yonemoto & Terasaki (226) administered thoracic duct lymphocytes obtained from HL-A–matched siblings to tumor patients (V). They reported two objective responses in six patients. An interesting technique was described by Symes & Riddell (227) in which pigs were sensitized to human tumors, then the pig lymphocytes were given to the tumor patient via the arterial distribution of the tumor (V). They obtained four objective responses in 14 patients. Vetto and colleagues (57) used transfer factor obtained from the cells of household contacts of tumor patients to treat advanced disease (V). Six objective responses were obtained in 35 patients; three of these responses were complete. Reinfusion of PHA-stimulated, autologous lymphocytes caused regression of pulmonary metastases in three of five patients with widespread metastatic tumor (VI) (228). However, metastases in other locations progressed and all of the patients succumbed to their disease.

While these studies represent unique approaches to the treatment of advanced tumors, they are logistically very difficult. Immunotherapy is of limited potency in dealing with advanced disease, and as the only treatment against established tumor, these approaches have been of low yield. However, they suggest possibilities that may be of value in combination with other treatment.

SUMMARY AND CONCLUSIONS

The bulk of clinical evidence indicates that immunotherapy is not a panacea for malignant disease, but then neither are other treatments now available. Because of its limited potency, immunotherapy is unlikely to become the primary treatment for any neoplasm except for those tumors accessible to topical or intralesional therapy. Immunotherapy depends on stimulation or augmentation of host defenses, which an established tumor has already overcome. Immunotherapy potentially can restore host defenses to a state that might have prevented the tumor in the first place, but it is not reasonable to expect more. The host has a capacity to deal with a limited number of cancer cells, but against gross established disease its defenses are limited.

Other treatments for cancer have limitations as well. Operation and radiation can control local disease, but fail to achieve cure because of microscopic systemic spread. Chemotherapy is potent and systemic in its effects, but kills tumor cells by first-order kinetics and may fail to eliminate the last tumor cell. Very little residual tumor is required to produce recurrence and death from most cancers. All of these treatment modalities have toxicities that occasion restraint in their use. Recognition of these limitations has resulted in a multimodality approach to cancer therapy, and in this setting a treatment that could selectively attack a small volume of tumor cells on a systemic basis would be very useful. The question at hand is whether immunotherapy can fill this role.

The most successful use of immunotherapy for malignant disease has been the intralesional injection of various agents to induce local DCH responses. Excellent control of local disease has been obtained in skin cancers and cutaneous metastases of malignant melanoma and breast carcinoma. However, it is not clear that this treatment improves survival, except for those long-term survivors who obtained dramatic control of regional malignant melanoma with BCG injections. Unfortunately, most patients go on to die of their disease. Whether there is destruction of systemic micrometastases, hence improved survival time following intralesional injection, has great import for the role of immunotherapy in cancer management. In experimental animals, intralesional injection of certain tumors confers systemic protection against subsequent tumor inocula (26). Following injection of melanoma in humans, there is a rise in antimelanoma antibody titer, and in some cases concomitant regression of uninjected lesions (12). This result indicates development of systemic resistance to tumor, but whether long-term control ensues is unanswered in human studies.

A direct extension of the question of improved survival time following intralesional injection is whether immunotherapy can be used to control microscopic residual disease following primary treatment for neoplasia. This is potentially the most useful area for immunotherapy, and there are several good, controlled studies outlined in this review that have shown benefit from such an approach. The most promising of these is the intrapleural injection of BCG following resection for early lung carcinoma (137, 140). Similary, repeated intradermal BCG treatment may benefit melanoma patients following resection of regional lymph node disease (45, 73). Levamisole seemed to help patients following radiation for advanced breast carcinoma and may have helped some patients after resection for lung carcinoma (39, 40). These studies are unconfirmed, but there are enough positive reports to lend credence to the hypothesis that immunotherapy results in systemic control of disease. It clearly warrants further investigation. A logical extension of adjuvant immunotherapy is to inject accessible tumors prior to surgical resection. This is currently under investigation for malignant melanoma and lung carcinoma.

Immunotherapy has been given in combination with chemotherapy for advanced malignant disease in an effort to increase the response and survival rates. For advanced breast cancer, both BCG and *C. parvum* have been added to chemotherapy with reportedly beneficial results (33, 73). As yet, these reports are unconfirmed, but no negative studies have appeared. In some controlled trials for acute myelogenous leukemia, the addition of immunotherapy to maintenance chemotherapy reportedly improves duration of remission and survival, but others have obtained negative results in the same setting. For disseminated melanoma, some investigators have claimed benefit for BCG added to chemotherapy, but there are several negative reports. The addition of MER or BCG to chemotherapy clearly does not improve results for advanced colon carcinoma. Overall, it seems that if chemotherapy produces a good response, which implies a marked lowering of tumor burden, immunotherapy seems to be additive. Otherwise, no effect is gained from the immunotherapy.

As the sole treatment for established tumors, immunotherapy plays a limited role. It may be considered if other alternatives have been exhausted. Some procedures have produced a few dramatic responses. A good example is cross transplantation of tumor and leukocyte exchange in patients with advanced melanoma. This has a fairly consistent response rate of about 15–20%, and has resulted in a few long-term regressions. This procedure involves the risk of severe transfusion reactions, however. These procedures are logistically difficult, potentially toxic, and help relatively few patients. However, these investigations do indicate that immunotherapy is not totally without effect against established disease.

In combination with other treatment modalities, immunotherapy appears to be effective against several types of human cancer, if the other modalities are able to reduce tumor burden to a low level. Any treatment that substantially lowers tumor burden is immunotherapy in the sense that it gives host defense mechanisms the opportunity to gain control of the disease. Immunotherapy is still an experimental treatment for cancer. However, there are enough good results from the clinical trials covered in this review to warrant continued investigation. There are many results, which, if confirmed in subsequent investigation, offer considerable benefit to selected patients.

There is much to be learned about the practical application of immunotherapy. The optimal dose and timing, as well as the route of administration, for most immunotherapeutic agents have not been fully determined. The short-term side effects have been documented, but long-term toxicities still must be defined. Enhancement of tumor growth has been seen in animal studies, and, while not conclusively demonstrated in humans, it is a potential problem. The brain is an immunologically privileged site and is where recurrence often appears. All of these problems have yet to be solved.

It would be helpful if these problems could be resolved in animal models. However, certain discrepancies exist between animal models and human cancer (229). Experimental animals are usually young, healthy, inbred adults. In contrast, the human tumor host is frequently very young or very old, and the population is genetically diverse. Experimental tumors are usually artificially induced, have a short latent period, rapid growth, and a relatively high immunogenicity. By comparison, human tumors arise spontaneously, have a long latent period, relatively slow growth, and a low immunogenicity. Animal models that more closely approximate the variability and unpredictability of human cancer must be developed. Because of the desperate nature of the disease, clinical investigators have been unwilling to wait for the development of these animal models before using a modality that is potentially beneficial for the cancer patient. At present, the well-conducted, carefully controlled, clinical trial is the major investigative tool for immunotherapy of human malignant disease.

Can the potency of immunotherapy be increased? The answer can only come from a full understanding of the mechanisms of tumor immunity and host defense against cancer. A variety of effector cells, and in some cases antibody, show activity against tumors in various assays. Knowing which effectors are the most important in vivo for tumor destruction, how they work, and what stimulates their production, might allow selective stimulation and augmentation of the most desirable response. Factors such as soluble tumor antigens and antigen-antibody complexes have been shown to retard host resistance to tumors. Finding ways to eliminate these adverse activities may result in a stronger antitumor response.

The development of accurate monitors of the effects of immunotherapy is an absolute necessity. As it stands now, the measure of effectiveness of immunotherapy in most situations is the maintenance of a disease-free state or increased survival time. Some patients do not benefit from the treatment or do not need it. There is a need for a test to identify the effects of immunotherapy, so that therapy that is not benefiting patients could be stopped or changed to an alternate form. Good monitoring would allow more rapid progress. An important side benefit from this research could be the ability to detect and follow early tumor growth.

Immunotherapy almost certainly has not reached its full potential, and many improvements are feasible. The results of these initial years of clinical investigation suggest that immunotherapy will have a role in the treatment of malignant disease. We are as Shakespeare's soothsayer described himself: "In nature's infinite book of secrecy, a little I can read." We only partially perceive the complex events involved in host resistance to cancer. Assuredly, progress will come as we learn to read more.

ACKNOWLEDGMENTS

This study was supported by USPHS grants CA12582, CA09010 awarded by the National Cancer Institute, DHEW: and the Medical Research Service of the Veterans Administration.

Literature Cited

1. Everson, T. C., Cole, W. H. 1966. *Spontaneous Regression of Cancer,* Philadelphia: Saunders
2. Foley, E. J. 1953. Antigenic properties of methylcholanthrene-induced tumors in mice of the strain of origin. *Cancer Res.* 13:835–37
3. Prehn, R. T., Main, J. M. 1957. Immunity to Methylcholanthrene-induced sarcomas. *J. Natl. Cancer Inst.* 18:769–78
4. Old, L. J., Boyse, E. A., Clark, D. A., Carswell, E. A. 1962. Antigenic properties of chemically-induced tumors. *Ann. NY Acad. Sci.* 101:80–106
5. Klein, G., Clifford, P., Klein, E., Stjernsward, J. 1966. Search for tumor-specific immune reactions in Burkitt lymphoma patients by the membrane immunoflourescence reaction. *Proc. Natl. Acad. Sci. USA* 55:1628–35
6. Hellstrom, I., Hellstrom, K. E., Pierce, G. E., Bill, A. H. 1968. Demonstration of cell bound and humoral immunity against neuroblastoma cells. *Proc. Natl. Acad. Sci. USA* 60:1231–38
7. Morton, D. L., Malmgren, R. A., Holmes, E. C., Ketcham, A. S. 1968. Demonstration of antibodies against human malignant melanoma by immunofluorescence. *Surgery* 64:233–40
8. Windhorst, D. 1976. The International Registry of Tumor Immunotherapy. *Med. Clin. North Am.* 60:641–48
9. Nauts, H. C., Fowler, G. A., Bogatko, F. H. 1953. A review of the influence of bacterial infection and of bacterial products (Coley's toxins) on malignant tumours in man. *Acta Med. Scand.* S276:1–103
10. Klein, G., Sjogren, H. O., Klein, E., Hellstrom, K. E. 1960. Demonstration of resistance against methylcholanthrene-induced sarcomas in the primary autochthonous host. *Cancer Res.* 20:1561–72
11. Mathe, G., Amiel, J. L., Schwarzenberg, L., Schneider, M., Cattan, A., Schlumberger, J. R., Hayat, M., De Vassal, F. 1969. Active immunotherapy for acute lymphoblastic leukemia. *Lancet* 1:697–99
12. Morton, D. L., Eilber, F. R., Malmgren, R. A., Wood, W. C. 1970. Immunological factors which influence response to immunotherapy in malignant melanoma. *Surgery* 68:158–64
13. Klein, E. 1968. Tumors of the skin. X. Immunotherapy of cutaneous and mucosal neoplasms. *NY State J. Med.* 68:900–11
14. Southam, C. M. 1967. Evidence for cancer-specific antigens in man. *Prog. Exp. Tumor Res.* 9:1–39
15. Hibbs, J. B., Lambert, L. J., Remington, J. S. 1972. Possible role of macrophage mediated nonspecific cytotoxicity in tumour resistance. *Nature New Biol.* 235:48–50
16. Alexander, P. 1976. The functions of the macrophage in malignant disease. *Ann. Rev. Med.* 27:207–24
17. Perlmann, P., Perlmann, H., Wigzell, H. 1972. Lymphocyte-mediated cytotoxicity *in vitro:* induction and inhibition by humoral antibodies and nature of effector cells. *Transplant Rev.* 13:91–116
18. Bast, R. C., Zbar, B., Borsos, T., Rapp, H. J. 1974. BCG and cancer (first of two parts). *N. Engl. J. Med.* 290:1413–20
19. Bast, R. C., Zbar, B., Borsos, T.,

Rapp, H. J. 1974. BCG and cancer (second of two parts). *N. Engl. J. Med.* 290:1458–69

20. Pilch, Y. H., Meyers, G. H. Jr., Sparks, F. C., Golub, S. H. 1975. In *Current Problems in Surgery,* ed. M. M. Ravitch, pp. 1–46, Chicago: Yearb. Med.
21. Golub, S. H. 1975. In *Cancer–A Comprehensive Treatise,* ed. F. F. Becker, Vol. 4, pp. 259–300. New York: Plenum
22. Zbar, B., Ribi, E., Kelly, M., Granger, D., Evans, C., Rapp, H. J. 1976. Immunologic approaches to the treatment of human cancer based on a guinea pig model. *Cancer Immunol. Immunother.* 1:127–37
23. Calmette, A. 1936. In *L'Infection Bacillaire et al Tuberculose Chez l'Homme et Chez l'Animaux,* p. 752. Paris: Masson
24. Nathanson, L., 1974. Use of BCG in the treatment of human neoplasms: a review. *Semin. Oncol.* 1:337–50
25. Klein, E., Klein, G. 1972. Specificity of homograft rejection *in vivo* assessed by inoculation of artificially mixed compatible and incompatible tumor cells. *Cell. Immunol.* 5:201–8
26. Hanna, M. G. Jr. 1974. Immunologic aspects of BCG mediated regression of established tumors and metastases in guinea pigs. *Semin. Oncol.* 1:319–35
27. Sparks, F. C., Silverstein, M. J., Hunt, J. S., Haskell, C. M., Pilch, Y. H., Morton, D. L. 1973. Complications of BCG immunotherapy in patients with cancer. *N. Engl. J. Med.* 289:827–30
28. Halpern, B. M., Biozzi, G., Stiffel, C., Mouton, D. 1966. Inhibition of tumor growth by administration of killed *Corynebacterium parvum. Nature New Biol.* 212:853–54
29. Woodruff, M. F. A., Boak, J. L. 1966. Inhibitory effect of injection of *Corynebacterium parvum* on the growth of tumour transplants in isogenic hosts. *Br. J. Cancer* 20:345–55
30. Milas, L., Hunter, N., Basic, I., Withers, H. R. 1974. Complete regressions of an established murine fibrosarcoma induced by systemic application of *Corynebacterium granulosum. Cancer Res.* 34:2470–75
31. Woodruff, M. F. A., Dunbar, N., Ghaffar, A. 1973. The growth of tumors in T-cell deprived mice and their response to treatment with *Corynebacterium parvum. Proc. R. Soc. London Ser. B.* 184:97–102
32. Oettgen, H. F., Pinsky, C. M., Delmonte, L. 1976. Treatment of cancer with immunomodulators. *Med. Clin. North Am.* 60:511–37
33. Israel, L., Edelstein, R. 1975. In *Immunological Aspects of Neoplasia,* pp. 485–504. Baltimore: Williams & Wilkins
34. Fisher, B., Rubin, H., Sartiano, G., Ennis, L., Wolmark, N. 1976. Observations following *Corynebacterium parvum* administration to patients with advanced malignancy: a phase I study. *Cancer* 38:119–30
35. Renoux, G., Renoux, M. 1973. Stimulation of anti-Brucella vaccination in mice by tetramisole, a phenylimidothiazole salt. *Infect. Immun.* 8:544–48
36. Tripodi, D., Parks, L. C., Brugmans, J. 1973. Drug-induced restoration of cutaneous delayed hypersensitivity in anergic patients with cancer. *N. Engl. J. Med.* 289:354–57
37. Brugmans, J., Schuermans, V., De Cock, W., Theinpont, D., Janssen, P., Verhaegen, H., Van Nimmen, L., Louwagie, A. C., Steven, E. 1973. Restoration of host defense mechanisms in man by levamisole. *Life Sci.* 13:1499–1504
38. Hoebeke, J., Franchi, G. 1973. Influence of tetramisole and its optical isomers on the mononuclear phagtocytic system. Effect on carbon clearance in mice. *J. Reticuloendothel. Soc.* 14:317–23
39. Rojas, A. F., Feierstein, J. N., Mickiewicz, E., Glait, H., Olivari, A. J. 1976. Levamisole in advanced human breast cancer. *Lancet* 1:211–15
40. Amery, W., Swierenga, J., Gooszen, H. C., Vanderschueren, R. G., Cosermans, J., Louwagie, A., Stam, J., Veldhuizen, R. W., Lopes Cordozo, E. 1975. Immunopotentiation with levamisole in resectable bronchogenic carcinoma: a double-blind controlled trial. *Br. Med. J.* 3:461–64
41. Ruuskanen, O., Remes, M., Makela, A. L., Isomaki, H., Toivanen, A. 1976. Levamisole and agranulocytosis. *Lancet* 2:958–59

42. Holland, J. F., Bekesi, J. G. 1976. Immunotherapy of human leukemia with neuraminidase-modified cells. *Med. Clin. North Am.* 60:539–49
43. Gillette, R. W., Boone, C. W. 1976. Augmented immunogenicity of tumor cell membranes produced by surface budding viruses: parameters of optimal immunization. *Int. J. Cancer* 18:216–22
44. Takita, H., Brugarolas, A. 1973. Adjuvant immunotherapy for bronchogenic carcinoma: preliminary results. *Cancer Chemother. Rep. Part 3* 4:293–98
45. Eilber, F. R., Morton, D. L., Holmes, E. C., Sparks, F. C., Ramming, K. P. 1976. Adjuvant immunotherapy with BCG in treatment of regional lymph node metastases from malignant melanoma. *N. Engl. J. Med.* 294:237–40
46. Sparks, F. C., Wile, A. G., Ramming, K. P., Silver, H. K. B., Wolk, R. W., Morton, D. L. 1976. Immunology and adjuvant chemoimmunotherapy of breast cancer. *Arch. Surg.* 111:1057–62
47. Townsend, C. M. Jr., Eilber, F. R., Morton, D. L. 1976. Skeletal and soft tissue sarcomas: treatment with adjuvant immunotherapy. *J. Am. Med. Assoc.* 236:2187–89
48. Morton, D. L. 1973. Horizons in tumor immunology.*Surgery* 74:69–79
49. Sumner, W. C., Foraker, A. G. 1960. Spontaneous regression of human melanoma: clinical and experimental studies. *Cancer* 13:79–81
50. Nadler, S. H., Moore, G. E. 1969. Immunotherapy of malignant disease. *Arch. Surg.* 99:376–81
51. Krementz, E. T., Mansell, P. W. A., Hornung, M. O., Samuels, M. S., Sutherland, C. A., Benes, E. N. 1974. Immunotherapy of malignant disease: the use of viable sensitized lymphocytes or transfer factor prepared from sensitized lymphocytes. *Cancer* 33:394–401
52. Jewell, W. R., Thomas, J. H., Sterchi, J. M., Morse, P. A., Humphrey, L. J. 1976. Critical analysis of treatment of stage II and stage III melanoma patients with immunotherapy. *Ann. Surg.* 183:543–49
53. Wright, P. W., Hellstrom, K. E., Hellstrom, I., Bernstein, I. D. 1976. Serotherapy of malignant disease. *Med. Clin. North Am.* 60: 607–22
54. Lawrence, H. S., 1969. Transfer factor. *Adv. Immunol.* 11:195–266
55. Levin, A. S., Byers, V. S., Fudenberg, H. H., Wybran, J., Hackett, A. J., Johnston, J. O., Spitler, L. E. 1975. Osteogenic sarcoma: immunologic parameters before and during immunotherapy with tumor-specific transfer factor. *J. Clin. Invest.* 55:487–99
56. LoBuglio, A. F., Neidhart, J. A. 1976. Transfer factor: a potential agent for cancer therapy. *Med. Clin. North Am.* 60:585–90
57. Vetto, R. M., Burger, D. R., Nolte, J. E., Vandenbark, A. A., Baker, H. W. 1976. Transfer factor therapy in patients with cancer. *Cancer* 37:90–97
58. Mannick, J. A., Egdahl, R. H. 1962. Transformation of nonimmune lymph node cells to state of transplantation immunity by RNA: a preliminary report. *Ann. Surg.* 156:356–66
59. Pilch, Y. H., Fritze, D., Kern, D. 1976. Immune RNA in the immunotherapy of cancer. *Med. Clin. North Am.* 60:567–83
60. Ramming, K. P., deKernion, J. B., Pilch, Y. H. 1977. In *Handbook of Cancer and Immunology,* ed. H. Waters, Washington: Smithsonian Inst. Press. In press
61. Trouillas, P., Lapras, C. 1969. Cellular immunotherapy of cerebral neurospongioma. *J. Med. Lyon* 1172:1269–91
62. Bloom, H. J. G., Peckham, M. J., Richardson, A. E., Alexander, P. A., Payne, P. M. 1973. Glioblastoma multiforme: A controlled trial to assess the value of specific active immunotherapy in patients treated by radical surgery and radiotherapy. *Br. J. Cancer* 27:253–67
63. Smith, G. V., Morse, P. A., Deraps, G. D., Raju, S., Hardy, J. D. 1973. Immunotherapy of patients with cancer. *Surgery* 74:59–68
64. Klein, E., Holtermann, O., Milgrom, H., Case, R. W., Klein, D., Rosner, D., Djerassi, I. 1976. Immunotherapy for accessible tumors utilizing delayed hypersensitivity reactions and separated components of the immune system. *Med. Clin. North Am.* 60:389–418
65. Garas, J., Besbeas, S., Papamatheakis, J., Gropas, G., Maragoudakis, S., Katsenis, A.,

Kiparissiadis, P., Konstadakos, P., Georgaka, A. 1975. Attempt with immunotherapy to control metastatic skin nodules from breast cancer by BCG. *Panminerva Med.* 17:193–95

66. Goldman, L. I. 1975. Immunotherapy of solid tumors. I. Preliminary studies with nitrogen mustard for nonspecific immunopotentiation in human cancer. *J. Surg. Res.* 18: 513–21
67. Stjernsward, J., Levin, A. 1971. Delayed hypersensitivity-induced regression of human neoplasms. *Cancer* 28: 628–40
68. Cheema, A. R., Hersh, E. M. 1972. Local tumor immunotherapy with *in vitro* autochthonous lymphocytes. *Cancer* 29:982–86
69. Copeland, M., ed. 1973. *Clinical staging system for carcinoma of the breast.* Chicago: Am. J. Comm. Cancer Staging End Results Rep., September
70. Anderson, J. M., Kelly, F., Wood, S. E., Halnan, K. E. 1974. Stimulatory immunotherapy in mammary cancer. *Br. J. Surg.* 61:778–84
71. Bonadonna, G., Brusamolino, E., Valagussa, P., Rossi, A., Brugnatelli, L., Brambilla, C., De Lena, M., Tancini, G., Bajetta, E., Musumeci, R., Veronesi, U. 1976. Combination chemotherapy as an adjuvant treatment in operable breast cancer. *N. Engl. J. Med.* 294:405–10
72. Pinsky, C. M., DeJager, R. L., Kaufman, R. J., Mike, V., Hansen, J. A., Oettgen, H. F., Krakoff, I. H. 1977. In *Immunotherapy of Cancer: Present Status of Trials in Man,* ed. W. D. Terry, D. Windhorst, New York: Raven. In press
73. Gutterman, J. U., Mavligit, G. M., Burgess, M. A., Cardenas, J. O., Blumenschein, G. R., Gottlieb, J. A., McBride, C. M., McCredie, K. B., Bodey, G. P., Rodriguez, V., Freireich, E. J., Hersh, E. M. 1976. Immunotherapy of breast cancer, malignant melanoma, and acute leukemia with BCG: prolongation of disease-free interval and survival. *Cancer Immunol. Immunother.* 1:99–107
74. Oettgen, H. F., Old, L. J., Farrow, J. H., Valentine, F. T., Lawrence, H. S., Thomas, L. 1974. Effects of dialyzable transfer factor in patients with breast cancer. *Proc. Natl. Acad. Sci. USA* 71:2319–23
75. Storer, E. H., Lockwood, R. A. 1969. In *Principles of Surgery,* ed. S. I. Schwartz, p. 958. New York: McGraw-Hill
76. Mavligit, G. M., Burgess, M. A., Seibert, G. B., Jubert, A. V., McBride, C. M., Gehan, E. A., Gutterman, J. U., Khankhanian, N., Speer, J. F., Martin, R. C., Copeland, E. M., Hersh, E. M. 1976. Prolongation of postoperative disease-free interval and survival in human colorectal cancer by BCG or BCG plus 5-fluorouracil. *Lancet* 1:871–75
77. Moertel, C. G., O'Connell, M. J., Ritts, R. E. Jr., Schutt, A. J., Reitemeier, R. J., Hahn, R. G., Frytak, S. K. 1977. See Ref. 72
78. Moertel, C. G., Ritts, R. E. Jr., Schutt, A. J., Hahn, R. G. 1975. Clinical studies of methanol extraction residue fraction of Bacillus Calmette-Guerin as an immunostimulant in patients with advanced cancer. *Cancer Res.* 35:3075–83
79. O'Connell, M. J., Ritts, R. E. Jr., Moertel, C. G. 1976. Immunological assessment of MER and placebo in advanced cancer: a double blind study. *Proc. Am. Assoc. Cancer Res.* 17:214
80. Engstrom, P. F., Paul, A. R., Catalano, R. B., Mastrangelo, M. J., Creech, R. H. 1977. See Ref. 72
81. Griffen, W. O. Jr., Meeker, W. R. 1972. Colon carcinoma and immunologic phenomena. *Surg. Clin. North Am.* 52:839–45
82. Falk, R. E., MacGregor, A. B., Landi, S., Ambus, U., Langer, B. 1976. Immunostimulation with intraperitoneally administered Bacillus Calmette Guerin for advanced malignant tumors of the gastrointestinal tract. *Surg. Gynecol. Obstet.* 142:363–68
83. Turowski, G., Cybulski, L., Politowski, M., Turaszwili, T., Zubel, M. 1976. First trials of immunopotentiation by thymus extract (TFX) in surgical patients with malignant disease. *Acta Med. Pol.* 17:19–39
84. Guinan, P., Bush, I. M., John, T., Sadoughi, N., Ablin, R. J. 1973. BCG immunotherapy in carcinoma of the prostate. *Lancet* 2:443–44
85. Merrin, C., Han, T., Klein, E., Wajsman, Z., Murphy, G. P. 1975. Immunotherapy of prostatic car-

cinoma with Bacillus Calmette Guerin. *Cancer Chemother. Rep.* 59:157–63

86. Eidinger, D., Morales, A. 1976. Treatment of superficial bladder cancer in man. *Ann. NY Acad. Sci.* 277:239–40
87. Ramming, K. P., deKernion, J. B. 1977. Immune RNA therapy for renal cell carcinoma: survival and immunologic monitoring. *Ann. Surg.* In press
88. Skinner, D. G., deKernion, J. B., Brower, P. A., Ramming, K. P., Pilch, Y. H. 1976. Advanced renal cell carcinoma: treatment with xenogeneic immune ribonucleic acid and appropriate surgical resection. *J. Urol.* 115:246–50
89. Guthrie, D., Way, S. 1975. Immunotherapy of nonclinical vaginal cancer. *Lancet* 2:1242–43
90. Graham, J. B., Graham, R. M. 1962. Autogenous vaccine in cancer patients. *Surg. Gynecol. Obstet.* 114:1–4
91. Imperato, S., Rossi, R., Ermiglia, G., De Marini, M., Cassolino, A. 1974. Active specific and nonspecific immunotherapy with immunological monitoring in late stage ovarian cancers. *Acta Eur. Fertil.* 5:25–39
92. Pattillo, R. A. 1976. Immunotherapy and chemotherapy of gynecologic cancers. *Am. J. Obstet. Gynecol.* 124:808–17
93. Rao, B., Wanebo, H. J., Ochoa, M., Lewis, J. L., Oettgen, H. F. 1977. Intravenous *corynebacterium parvum*: an adjunct to chemotherapy for resistant advanced ovarian carcinoma. *Cancer* 39:514–26
94. Hudson, C. N., McHardy, J. E., Curling, O. M., English, P. E., Levin, L., Poulton, T. A., Crowther, M., Leighton, M. 1976. Active specific immunotherapy for ovarian cancer. *Lancet* 2:877–79
95. Mathé, G., Schwarzenberg, L., De Vassal, F., Delgado, M. 1976. Immunotherapy for acute lymphoid leukemia. *Lancet* 1:143–44
96. Otten, J. 1977. See Ref. 72
97. Mathé, G., Pouillart, P., Schwarzenberg, L., Amiel, J. L., Schneider, M., Hayat, M., De Vassal, F., Jasmin, C., Rosenfeld, C., Weiner, R., Rappaport, H. 1972. Attempts at immunotherapy of 100 patients with acute lymphoid leukemia: some factors influencing results. *Natl. Cancer Inst. Monogr.* 35:361–71
98. Medical Research Council. 1971. Treatment of acute lymphoblastic leukaemia: comparison of immunotherapy (BCG), intermittent methotrexate, and no therapy after a five-month intensive cytotoxic regimen (Concord trial). *Br. Med. J.* 4:189–94
99. Heyn, R. M., Joo, P., Karon, M., Nesbit, M., Shore, N., Breslow, N., Weiner, J., Reed, A., Hammond, D. 1975. BCG in the treatment of acute lymphocytic leukemia. *Blood* 46:431–42
100. Pinkel, D., Simone, J., Hustu, H. O., Rhomes, J. A. A. 1972. Nine years' experience with "total therapy" of childhood acute lymphocytic leukemia. *Pediatrics* 50: 246–51
101. Poplack, D. G., Graw, R. G., Pomeroy, T. C., Henderson, E. S., Leventhal, B. G. 1975. Chemotherapy (CT) vs. chemotherapy and immunotherapy (CT + IMT) in childhood acute lymphatic leukemia (ALL). *Proc. Am. Soc. Clin. Oncol.* 16:230
102. Leventhal, B. G., LePourhiet, A., Halterman, R. H., Henderson, E. S., Herberman, R. B. 1973. Immunotherapy in previously treated acute lymphatic leukemia. *Natl. Cancer Inst. Monogr.* 39:177–87
103. Sacks, K. L., Olweny, C., Mann, D. L., Simon, R., Johnson, G. E., Poplack, D. G., Leventhal, B. G. 1975. A clinical trial of chemotherapy and RAJI immunotherapy in advanced acute lymphatic leukemia. *Cancer Res.* 35:3715–20
104. Ekert, H., Jose, D. G. 1975. Chemotherapy and BCG in acute lymphocytic leukemia. *Lancet* 2: 713–14

104a. Kay, H. 1977. See Ref. 72

105. Powles, R. L., Crowther, D., Bateman, C. J. T., Beard, M. E. J., McElwain, T. J., Russel, J., Lister, T. A., Whitehouse, J. M. A., Wrigley, P. F. M., Pike, M., Alexander, P., Hamilton-Fairley, G. 1973. Immunotherapy for acute myelogenous leukaemia. *Br. J. Cancer* 28:365–76
106. Vogler, W. R., Chan, Y. K. 1974. Prolonging remission in myeloblastic leukemia by tice strain Bacillus Calmette Guerin. *Lancet* 2:128–31
107. Gutterman, J. U., Rodriguez, V.,

Mavligit, G. M., Burgess, M. A., Gehan, E., Hersh, E. M., McCredie, K. B., Reed, R., Smith, T., Bodey, G. P., Freireich, E. J. 1974. Chemoimmunotherapy of adult acute leukaemia: prolongation of remission in myeloblastic leukaemia with BCG. *Lancet* 2:1405–9

108. Peto, R., Galton, D. A. G. 1975. Chemoimmunotherapy of adult acute leukaemia. *Lancet* 1:454
109. Gutterman, J. U., Hersh, E. M., Gehan, E. A., Freireich, E. J. 1975. Chemoimmunotherapy of adult acute leukaemia. *Lancet* 1:454–55
110. Whiteside, M. G., Cauchi, M. N., Paton, C. Stone, J. 1976. Chemoimmunotherapy for maintenance in acute myeloblastic leukemia. *Cancer* 38:1581–86
111. Reizenstein, P., Brenning, G., Engstedt, L., Franzen, S., Gahrton, G., Gullbring, B., Holm, G., Hocker, P., Hoglund, S., Hornsten, P., Jameson, S., Killander, A., Killander, D., Klein, E., Lantz, B., Lindemalm, C., Lockner, D., Lonnqvist, B., Mellstedt, H., Palmblad, J., Pauli, C., Skarberg, K. O., Uden, A. M., Vanky, F., Wadman, B. 1977. See Ref. 72
112. Peto, R. 1977. See Ref. 72
113. Hewlett, J. S., Balcerzak, S., Gutterman, J. 1977. See Ref. 72
114. Freeman, C. B., Harris, R. G., Colin, G., Leyland, M. J., MaCiver, J. E., Delamore, I. W. 1973. Active immunotherapy used alone for the maintenance of patients with acute myeloid leukaemia. *Br. Med. J.* 4:571–73
115. Powles, R. L. 1976. Immunologic maneuvers in the management of acute leukemia. *Med. Clin. North Am.* 60:463–72
116. Bodey, G. P., Freireich, E. J., Gehan, E. A., McCredie, K. B., Rodriguez, V., Gutterman, J. U., Burgess, M. A. 1976. Late intensification therapy for acute leukemia in remission. Chemotherapy and immunotherapy. *J. Am. Med. Assoc.* 235:1021–25
117. Whittaker, J. A., Slater, A. J. 1977. See Ref. 72
118. Cuttner, J., Glidewell, O., Holland, J. F. 1977. See Ref. 72
119. Schwarzenberg, L., Mathé, G., Schneider, M., Amiel, J. L., Cattan, A., Schlumberger, J. R. 1966. Attempted adoptive immunotherapy of acute leukaemia by leucocyte transfusions. *Lancet* 2:365–68
120. Gee, T. S., Clarkson, B. D. 1977. See Ref. 72
121. McIntyre, O. R., Rai, K., Glidewell, O., Holland, J. F. 1977. See Ref. 72
122. Sauter, C., Cavalli, F., Lindenmann, J., Gmur, J., Berchtold, W., Alberto, P., Obrecht, P., Senn, H. J. 1977. See Ref. 72
123. Aleksandrowicz, V., Skotnicki, A. B. 1976. The role of the thymus and thymic humoral factors in immunotherapy of aplastic and proliferation disease of the hemopoietic system. *Acta Med. Pol.* 17:1–17
124. DeCarvalho, S. 1963. Preliminary experimentation with specific immunotherapy of neoplastic disease in man. I. Immediate effects of hyperimmune equine gamma globulins. *Cancer* 16:306–30
125. Albo, V., Hartmann, J. R., Krevit, M. 1968. Fresh plasma as an adjuvant to the chemotherapy of previously treated children with ALL. *Proc. Am. Assoc. Clin. Res.* 9:2
126. Skurkovich, S. V., Makhonova, L. A., Reznichenko, F. M., Chervonskiy, G. I. 1969. Treatment of children with acute leukemia by passive cyclin immunization with autoplasma and autoleukocytes operated during the remission period. *Blood* 33:186–97
127. Sokal, J. E., Aungst, C. W., Grace, J. T. Jr. 1973. Immunotherapy in well-controlled chronic myelocytic leukemia. *NY State J. Med.* 73:1180–85
128. Laszlo, J., Buckley, C. I. III, Amos, D. B. 1968. Infusion of isologous immune plasma in chronic lymphocytic leukemia. *Blood* 31: 104–10
129. Klein, G. 1975. The Epstein-Barr virus and neoplasia. *N. Engl. J. Med.* 293:1353–57
130. Magrath, I. T., Ziegler, J. L. 1976. Failure of BCG immunostimulation to affect the clinical course of Burkitt's lymphoma. *Br. Med. J.* 1:615–18
131. Burkitt, D. 1967. In *Treatment of Burkitt's Tumour,* ed. J. H. Burchenal, D. Burkitt, p. 197. New York: Springer-Verlag
132. Ngu, V. A. 1967. See Ref. 131, p. 204
133. Fass, L., Herberman, R. B., Ziegler, J., Morrow, R. H. Jr. 1970. Evaluation of the effect of remission

plasma on untreated patients with Burkitt's lymphoma. *J. Natl. Cancer Inst.* 44:145–49

134. Bluming, A. Z., Serpick, A. A. 1970. Treatment of recurrent Burkitt's tumor with autologous remission plasma. *Oncology* 24:304–12
135. Sokal, J. E., Aungst, C. W., Snyderman, M. 1974. Delay in progression of malignant lymphoma after BCG vaccination. *N. Engl. J. Med.* 291:1226–30
136. Jones, S. E., Salmon, S. E., eds. 1977. *Adjuvant Therapy of Cancer,* pp. 549–56. Amsterdam: North-Holland
137. McKneally, M. F., Maver, C., Kausel, H. W. 1976. Regional immunotherapy of lung cancer with intrapleural BCG. *Lancet* 1:377–79
138. Takita, H. 1970. Effect of postoperative empyema on survival of patients with bronchogenic carcinoma. *J. Thorac. Cardiovasc. Surg.* 59:642–44
139. Ruckdeschel, J. C., Codish, S. D., Stranahan, A., McKneally, M. F. 1972. Postoperative empyema improves survival in lung cancer: documentation and analysis of a natural experiment. *N. Engl. J. Med.* 287:1013–17
140. McKneally, M. F., Maver, C. M., Kausel, H. W. 1977. Intrapleural BCG stimulation in lung cancer. *Lancet* 1:593
141. Edwards, F. R., Whitwell, F. 1974. Use of BCG as an immunostimulant in the surgical treatment of carcinoma of the lung. *Thorax* 29:654–58
142. Amery, W. K. 1977. See Ref. 72
143. Pines, A. 1976. A 5 year controlled study of BCG and radiotherapy for inoperable lung cancer. *Lancet* 1:380–81
144. Takita, H., Takada, M., Minowada, J., Han, T., Edgerton, F. 1977. See Ref. 72
145. Stewart, T. H. M., Hollinshead, A. C., Harris, J. E., Raman, S. 1977. See Ref. 72
146. Dimitrov, N. V., Conroy, J., Suhrland, L. G., Singh, T., Teitlebaum, H. 1977. See Ref. 72
147. Alth, G., Denck, H., Fischer, M., Karrer, K., Kokron, O., Korizek, E., Micksche, M., Ogris, E., Reider, C., Titscher, R., Wrba, H. 1973. Aspects of the immunologic treatment of lung cancer. *Cancer Chemother. Rep. Part. 3,* 4:271–74
148. Kimura, I., Ohnoshi, T., Yasuhara, S., Sugiyama, M., Urabe, Y., Fujii, M., Machida, K. I. 1976. Immunochemotherapy in human lung cancer using the streptococcal agent OK-432. *Cancer* 37:2201–3
149. Lieberman, R., Spitler, L. E., Wybran, J., Fudenberg, H. H., Levinson, D., Epstein, W. 1974. Clinical and immunologic evaluation in patients with metastatic malignant melanoma (cutaneous only) treated with intralesional BCG therapy. *Clin. Res.* 22:130A
150. Israel, L., Depierrre, A., Edelstein, R., Cros-Decan, J., Maury, P. 1975. Effect of intranodular BCG in 22 melanoma patients. *Panminerva Med.* 17:187–88
151. Mastrangelo, M. J., Sulit, H. L., Prehn, L. M., Bornstein, R. S., Yarbro, J. W., Prehn, R. T. 1976. Intralesional BCG in the treatment of metastatic malignant melanoma. *Cancer* 37:684–92
152. Gerner, R. E., Moore, G. E. 1976. Feasibility study of active immunotherapy in patients with solid tumors. *Cancer* 38:131–43
153. Morton, D. L., Eilber, F. R., Holmes, E. C., Hunt, J. S., Ketcham, A. S., Silverstein, M. J., Sparks, F. C. 1974. BCG immunotherapy of malignant melanoma: summary of a seven-year experience. *Ann. Surg.* 180:635–43
154. Karakousis, C. P., Douglass, H. O. Jr., Yeracaris, P. M., Holyoke, E. D. 1976. BCG immunotherapy in patients with malignant melanoma. *Arch. Surg.* 111:716–18
155. Yamamura, Y. 1977. See Ref. 72
156. Nathanson, L. 1972. Regression of intradermal malignant melanoma after intralesional injection of *Mycobacterium bovis* strain BCG. *Cancer Chemother. Rep.* 56:659–65
157. Seigler, H. F., Shingleton, W. W., Metzgar, R. S., Buckley, C. E., Bergoc, P. M., Miller, D. S., Fetter, B. G., Phaup, M. D. 1972. Nonspecific and specific immunotherapy in patients with melanoma. *Surgery* 72:162–74
158. Baker, M. A., Taub, R. N. 1973. BCG in malignant melanoma. *Lancet* 1:1117–18
159. Pinsky, C. M., Hirshaut, Y., Oettgen, H. F. 1973. Treatment of malignant melanoma by intratumoral injection of BCG. *Natl. Cancer Inst. Monogr.* 39:225–28
160. Grant, R. M., Cochran, A. J., Hoyle, D., Mackie, R., Murray,

E. L., Ross, C. 1974. Results of administering BCG to patients with melanoma. *Lancet* 2:1096–1100

161. Klein, E., Holtermann, O. A. 1972. Immunotherapeutic approaches to the management of neoplasms. *Natl. Cancer Inst. Monogr.* 35:379–91

162. Malek-Mansour, S. 1973. Remission of melanoma with DNCB treatment. *Lancet* 27:503–4

163. Burdick, K. H., Hawk, W. A. 1961. Vitiligo in a case of vaccinia virus-treated melanoma. *Cancer* 17:708–12

164. Milton, G. W., Lane Brown, M. M. 1966. The limited role of attenuated smallpox virus in the management of advanced malignant melanoma. *Aust. NZ J. Surg.* 35:286–90

165. Hunter-Craig, I., Newton, K. A., Westburg, G., Lacey, B. W. 1970. Use of vaccinia virus in the treatment of metastatic malignant melanoma. *Br. Med. J.* 2:512–15

166. Minton, J. P. 1973. Mumps virus and BCG vaccine in metastatic melanoma. *Arch. Surg.* 106:503–6

167. Cunningham-Rundles, W. F., Hirshaut, Y., Pinsky, C. M., Oettgen, H. F. 1975. Phase I trial of intralesional *Corynebacterium parvum*. *Clin. Res.* 23:337A

167a. Morton, D. L., Eilber, F. R., Holmes, E. C., Townsend, C. M., Mirra, J., Weisenburger, T. 1977. Adjuvant therapy in melanomas and sarcomas. See Ref. 136, pp. 391–97

168. Pinsky, C. M., Hirshaut, Y., Wanebo, H. J., Fortner, J., Mike, V., Schottenfeld, D., Oettgen, H. F. 1977. See Ref. 72

169. Eilber, F. R., Townsend, C. M. Jr., Morton, D. L. 1976. Results of BCG adjuvant immunotherapy for melanoma of the head and neck. *Am. J. Surg.* 132:476–79

170. Bluming, A. Z., Vogel, C. L., Ziegler, J. L., Mody, N., Kamya, G. 1972. Immunological effects of BCG in malignant melanoma: two modes of administration compared. *Ann. Intern. Med.* 76:405–11

171. Ikonopisov, R. L. 1972. The rationale of immunostimulation procedures in the therapeutic approach to malignant melanoma of the skin. *Tumori* 58:121–27

172. Gutterman, J. U., McBride, C. M., Freireich, E. J., Mavligit, G. M., Frei, E. III, Hersh, E. M. 1973. Active immunotherapy with BCG for recurrent malignant melanoma. *Lancet* 1:1208–12

173. Serrou, B., DuBois, J. B., Meiss, L., Romieu, C. 1975. Local active immunotherapy for malignant melanoma. *Panminerva Med.* 17:182–86

174. MacGregor, A. B., Falk, R. E., Landi, S., Ambus, U., Langer, B. 1975. Oral bacillus Calmette Guerin immunostimulation in malignant melanoma. *Surg. Gynecol. Obstet.* 141:747–54

175. Thompson, R. B. 1975. Active immunotherapy with BCG for melanoma. *Panminerva Med.* 17:189–92

176. Cunningham, T. J., Schoenfeld, D., Wolters, J., Nathanson, L., Cohen, M., Patterson, B. 1977. See Ref. 72

177. Beretta, G. 1977. See Ref. 72

178. Spitler, L. E. Sagebiel, R., Wong, P., Malm, T., Chase, R., Gonzalez, R. 1977. See Ref. 72

179. Gutterman, J. U., Mavligit, G. M., Gottlieb, J. A., Burgess, M. A., McBride, C. E., Einhorn, L., Freireich, E. J., Hersh, E. M. 1974. Chemoimmunotherapy of disseminated malignant melanoma with dimethyl-triazeno imidazole carboxamide and bacillus Calmette-Guerin. *N. Engl. J. Med.* 291:592–96

180. Currie, G. A., McElwain, T. J. 1975. Active immunotherapy as an adjunct to chemotherapy in the treatment of disseminated malignant melanoma: a pilot study. *Br. J. Cancer* 31:143–56

181. Newlands, E. S., Oon, C. J., Roberts, J. T., Elliott, P., Mould, R. F., Topham, C., Madden, F. J. F., Newton, K. A., Westburg, G. 1976. Clinical trial of combination chemotherapy and specific active immunotherapy in disseminated melanoma. *Br. J. Cancer* 34:174–79

182. Gutterman, J. U., Mavligit, G. M., Reed, R., Burgess, M. A., Gottlieb, J., Hersh, E. M. 1976. Bacillus Calmette Guerin immunotherapy in combination with DTIC (NSC-45388) for the treatment of malignant melanoma. *Cancer Treat. Rep.* 60:177–82

183. Costanzi, J. J. 1977. See Ref. 72

184. Mastrangelo, M. J., Bellet, R. E., Berd, D. 1977. See Ref. 72

185. Presant, C. A., Bartolucci, A., Smalley, R., Vogler, W. R. 1977. See Ref. 72

186. Curtis, J. E. 1971. Adoptive immunotherapy in the treatment of advanced malignant melanoma. *Proc. Am. Assoc. Clin. Res.* 12:52

187. Krementz, E. T., Samuels, M. S., Wallace, J. H., Benes, E. N. 1971. Clinical experiences in immunotherapy of cancer. *Surg. Gynecol. Obstet.* 133:209–17
188. Goodwin, D. P., Hornung, M. O., Krementz, E. T. 1973. Extraction and use of melanoma-associated protein for immunotherapy. *Oncology* 27:258–65
189. Spitler, L. E., Levin, A. S., Blois, M. S., Epstein, W., Fudenberg, H. H., Hellstrom, I., Hellstrom, K. E. 1972. Lymphocyte responses to tumor-specific antigens in patients with malignant melanoma and results of transfer factor therapy. *J. Clin. Invest.* 51:92
190. Brandes, L. J., Galton, D. A. G., Wiltshaw, E. 1971. New approach to immunotherapy of melanoma. *Lancet* 2:293–5
191. Morse, P. A. Jr., Deraps, G. D., Smith, G. V., Raju, S., Hardy, J. D. 1973. Transfer factor therapy of human cancer. *Clin. Res.* 21:71
192. Price, F. B., Hewlett, J. S., Deodhar, S. D., Barna, B. 1974. The therapy of malignant melanoma with transfer factor. *Cleveland Clin. Q.* 41:1–4
193. Seigler, H. F., Shingleton, W. W., Metzgar, R S., Buckley, C. E., Bergoc, P. M. 1973. Immunotherapy in patients with melanoma. *Ann. Surg.* 178:352–9
194. Levy, N. L., Seigler, H. F., Shingleton, W. W. 1974. A multiphase immunotherapy regimen for human melanoma: clinical and laboratory results. *Cancer* 34:1548–57
195. Ghose, T., Norvell, S. T., Guclu, A., Cameron, D., Bodurtha, A., MacDonald, A. S. 1972. Immunochemotherapy of cancer with chloramubucil-carrying antibody. *Br. Med. J.* 3:495–99
196. Falk, R. E., Mann, P., Langer, B. 1973. Cell-mediated immunity to human tumours. *Arch. Surg.* 107:261–65
197. Morton, D. L., Malmgren, R. A. 1968. Human osteosarcomas: immunologic evidence suggesting an associated infectious agent. *Science* 102:1279–81
198. Marcove, R. C., Southam, C. M., Levin, A., Mike, V., Huvos, A. 1971. A clinical trial of autogenous vaccine in osteogenic sarcoma in patients under the age of twenty-five. *Surg. Forum* 22:434–35
199. Marsh, B., Flynn, L., Enneking, W. J. 1972. Immunologic aspects of osteosarcoma and their application to therapy. *J. Bone Jt. Surg.* 54A: 1367–97
200. Neff, J. R., Enneking, W. F. 1975. Adoptive immunotherapy in primary osteosarcoma. *J. Bone Jt. Surg.* 57 A:145–48
201. Eilber, F. R., Townsend, C. M.Jr., Morton, D. L. 1975. Osteosarcoma: results of treatment employing adjuvant immunotherapy. *Clin. Orthop. Relat. Res.* 111:94–100
202. Taylor, W. F., Ivins, J. C., Dahlin, D. C., Pritchard, D. J. 1977. See Ref. 72
203. Ritts, R. E. Jr., Pritchard, D. J., Gilchrist, G. S., Ivins, J. C., Taylor, W. F. 1977. See Ref. 72
204. LoBuglio, A. F., Neidhart, J. A., Hilberg, R. W., Metz, E. N., Balcerzak, S. P. 1973. The effect of transfer factor therapy on tumor immunity in alveolar soft part sarcoma. *Cell. Immunol.* 7:159–65
205. Morton, D. L., Eilber, F. R., Joseph, W. L., Wood, W. C., Trahan, E., Ketcham, A. S. 1970. Immunological factors in human sarcomas and melanomas: a rational basis for immunotherapy. *Ann. Surg.* 172:740–49
206. Morton, D. L., Holmes, E. C., Eilber, F. R., Wood, W. C. 1971. Immunological aspects of neoplasia: a rational basis for immunotherapy. *Ann. Intern. Med.* 74:587–604
207. Sinkovics, J. G. 1977. Immunotherapy with viral oncolysates for sarcoma. *J. Am. Med. Assoc.* 237: 869
208. Green, A. A., Pratt, C., Webster, R. G., Smith, K. 1976. Immunotherapy of osteosarcoma patients with virus-modified tumor cells. *Ann. NY Acad. Sci.* 277:396–410
209. Gunnarson, A., McKhann, C. F., Simmons, R. L., Grage, T. B. 1974. Metastatic sarcoma: combined surgical and immunotherapeutic approach: neuraminidase treated tumor cells as tumor vaccine. *Minn. Med.* 57:558–61
210. Klein, E. 1969. Hypersensitivity reactions at tumor sites. *Cancer Res.* 29:2351–62
211. Levis, W. R., Kraemer, K. H., Klinger, W. G., Peck, G. L., Terry, W. D. 1973. Topical immunotherapy of basal cell car-

cinomas with dinitrochlorobenzene. *Cancer Res.* 33:3036–42

212. Williams, A. C., Klein, E. 1970. Experiences with local chemotherapy and immunotherapy in premalignant and malignant skin lesions. *Cancer* 25:450–62
213. Raaf, J. H., Krown, S. E., Pinsky, C. M., Cunningham-Rundles, W., Safai, B., Ottegen, H. F. 1976. Treatment of Bowen's disease with topical dinitrochlorobenzene and 5-fluorouracil. *Cancer* 37:1633–42
214. Villasor, R. P. 1965. The clinical use of BCG vaccine in stimulating host resistance to cancer. *J. Philipp. Med. Assoc.* 41:619–32
215. Asada, T. 1974. Treatment of human cancer with mumps virus. *Cancer* 34:1907–28
216. Israel, L., Edelstein, R. 1975. A phase II study of daily intravenous corynebacterium in 33 disseminated cancers. *Proc. Am. Assoc. Cancer Res.* 16:67
217. Aswaq, M., Richards, V., McFadden, S. 1964. Immunologic response to autologous cancer vaccine. *Arch. Surg.* 89:485–87
218. Hughes, L. E., Kearney, R., Tully, M. 1970. A study in clinical cancer immunotherapy. *Cancer* 26:269–78
219. Czajkowski, N. P., Rosenblatt, M., Wolf, P. L., Vasquez, J. 1967. A new method of active immunization to autologous human tumour tissue. *Lancet* 2:905–9
220. Cunningham, T. J., Olson, K. B., Laffin, R., Horton, J., Sullivan, J. 1969. Treatment of advanced cancer with active immunization. *Cancer* 24:932–37
221. Humphrey, L. J., Boehm, B., Jewell, W. R., Boehm, O. R. 1972. Immunologic response of cancer patients modified by immunization with tumor vaccine. *Ann. Surg.* 176:554–58
222. Wallack, M. K., Steplewski, Z., Koprowski, H., Rosato, E., George, J., Hulihan, B., Johnson, J. 1977. A new approach in specific active immunotherapy. *Cancer* 39:560–64
223. Finney, J. W., Byers, E. H., Wilson, R. H. 1960. Studies in tumor autoimmunity. *Cancer Res.* 20:351–56
224. Woodruff, M. F. A., Nolan, B. 1963. Preliminary observations on treatment of advanced cancer by injection of allogeneic spleen cells. *Lancet* 2:426–29
225. Andrews, G. A., Congdon, C. C., Edwards, C. L., Gengozian, N., Nelson, B., Vodopick, H. 1967. Preliminary trials of clinical immunotherapy. *Cancer Res.* 27:2535–41
226. Yonemoto, R. H., Terasaki, P. I. 1972. Cancer immunotherapy with HLA compatible thoracic duct lymphocyte transplantation. *Cancer* 30:1438–43
227. Symes, M. O., Riddell, A. G. 1973. The use of immunized pig lymph node cells in the treatment of patients with advanced malignant disease. *Br. J. Surg.* 60:176–80
228. Frenster, J. H. 1976. Phytohemagglutinin-activated autochthonous lymphocytes for systemic immunotherapy of human neoplasms. *Ann. NY Acad. Sci.* 277:45–50
229. Bartlett, G. L., Kreider, J. W., Purnell, D. M. 1976. Immunotherapy of cancer in animals: models or muddles? *J. Natl. Cancer Inst.* 56:207–10

Ann. Rev. Med. 1978. 29:285–306
Copyright © 1978 by Annual Reviews Inc. All rights reserved

OTITIS MEDIA: THE SPECTRUM OF MIDDLE EAR INFLAMMATION[1]

◆7280

G. Scott Giebink, M.D. and Paul G. Quie, M.D.
Departments of Pediatrics and Microbiology, University of Minnesota, School of Medicine, Minneapolis, Minnesota 55455

Otitis media is characterized by a broad range of pathophysiologic events occurring within the middle ear. Only recently have these events been defined by terminology acceptable to both clinicians and researchers (1). All middle ear effusions appear to result from inflammation of the middle ear cleft. Acute suppurative (purulent) otitis and chronic nonsuppurative (serous and mucoid) otitis have been described. Combinations, such as mucopurulent and seropurulent otitis, also occur. Prior to the development of sensitive diagnostic tools, such as the pneumatic otoscope and tympanometer, diagnosis of otitis media was based largely on the presence of an inflammed tympanic membrane and/or on the presence of otorrhea. Thus, until recently we have primarily detected acute otitis media only, and it is mostly from this experience that the natural history of otitis media has been studied. This review describes the (*a*) incidence, (*b*) pathogenesis, (*c*) diagnosis, (*d*) management, and (*e*) complications of acute and chronic otitis media.

INCIDENCE

The worldwide magnitude of otitis media, particularly among infants and young children, has been extrapolated from point-prevalence studies, studies of isolated populations, and recently from long-term prospective evaluation beginning at birth of children. This disease has been most completely studied with minimal therapeutic intervention among 489 Eskimo children surveyed periodically by Kaplan et al (2). Overall, 76% of the children had at least one episode of acute otitis media defined by the history of otorrhea; of these, 78% had their first episode before age two years. Forty-one percent of the children displayed either

---

[1] This work was supported in part by USPHS grants AI 06931 and AI 08821 and Contract #N01-AI-52533 from the NIAID, NIH.

0066-4219/78/0401-0285$01.00

chronic tympanic membrane perforation or significant scarring. Otitis media has also been studied and found to occur in high frequency among the Maori of New Zealand (3), the Aborigines of Australia (4), and American Indians, 50% of whom were found to have chronic perforations (5).

The prevalence of acute otitis media was first described in the population at large by the British Medical Research Council in 1957 (6). Of more than 47,500 individuals, 1162 (2.4%) had at least one episode of acute otitis media during a one year observation period. However, 10–20% of children below age eight years had an acute episode, and the risk of having further attacks increased with each successive episode. Brownlee et al (7) more carefully defined the natural history of acute otitis media among children in a large private practice group in South Carolina. Using retrospective analysis of 772 five year olds, 84% were noted to have had at least one bout of acute otitis media. Fifty percent of all children had three or more episodes, and half of these had at least six episodes. The peak incidence of otitis media (18%) was between ages one and two years, and remained above 10% through the sixth year. Thereafter, the incidence declined to less than 1% by the twelfth year.

Howie et al (8) prospectively observed a cohort of 488 children born into their private practice in Alabama. Otitis media was defined as the presence of middle ear effusion detected by pneumatic otoscopy. Forty-nine percent of these children were noted to have otitis media during their first year of life, and an additional 14% had their first bout during the second year (Table 1). After two years of age the incidence of first bouts remained between 4 and 5%.

Recurrent episodes were quite common among those patients who had a prior bout of otitis media (Table 1). If the initial episode occurred during the first 12 months of life, recurrence was especially common. Patients who had at least two episodes during their first year had at least twice as many recurrences as did those with one or no episodes of otitis media during the first year.

The question of disease prevalence is no doubt related to the difficulties in examining young children and in detecting middle ear effusion with or without acute inflammation. An example of these difficulties is found in infants with cleft palate. Until recently, effusions were considered uncommon in children and

Table 1 Incidence of otitis media in children followed from birth[a]

	Patients with otitis media (%)		
Age (yr)	First episode	Subsequent episodes	Total
0–1	49	—	49
1–2	14	27	41
2–3	4	21	25
3–4	5	18	23
4–5	4	14	18
5–6	4	18	22

[a] Adapted from Howie, V. M. et al. See Ref. 8.

adults with cleft palate (9). However, on close examination of affected infants, it has been apparent that their ears universally contain effusion (10, 11). Persistence of middle ear effusion, identified by tympanocentesis following acute otitis media, has also been described in healthy children without maxillofacial abnormalities (12). Of 93 children followed for at least four weeks after the acute episode, 37.5% had persistent effusion. Other studies employing audiometry to diagnose effusion, have shown that the risk of persistent effusion and hearing loss varies by age group and population studied, but is present in all populations. Thus, the epidemiologic data accumulated to date on the prevalence of otitis media, considering both its acute and chronic forms, most likely underestimates the magnitude of this disease.

PATHOGENESIS

Mechanical

The pathogenesis of otitis media is intimately related to mechanical ventilation of the middle ear cleft. Obstructed ventilation, whether produced by extrinsic blockage of the eustachian tube, by intrinsic inflammation, or by other immunochemical events, appears to result in middle ear effusion. Although many investigators have postulated that eustachian tube malfunction is the principal pathogenic mechanism for otitis media, satisfactory evidence of tubal pathophysiology is lacking. Normal eustachian tube function provides ventilation of the middle ear, drainage of middle ear secretions, and protection of the middle ear from severe acoustic noise. The physiology of the eustachian tube has been extensively studied in normal individuals and in those with middle ear effusion. Cantekin et al (13) found evidence for tubal malfunction in children with chronic forms of otitis media when compared to children with traumatic tympanic membrane perforation but without a history of otitis media. Employing similar techniques for measuring air pressure equalization via the eustachian tube, Holmquist & Renvall (14) also found abnormal tubal function in over 75% of patients with chronic otitis media up to 40 weeks after placement of tympanic membrane ventilating tubes. Unfortunately, the method employed by these investigators for studying tubal function requires a perforated tympanic membrane, which restricts, the application of this method. Ingelstedt (15) described a noninvasive method for measuring tubal physiology by recording the volume deviations of the tympanic membrane during induced changes in ambient and middle ear pressures. Study of 102 healthy adult volunteers showed that the pressure-equalizing ability of the tube was correlated with deglutition, yawning, Valsalva maneuver, and nose blowing; tubal function was reduced when patients were recumbent. Kirchner et al (16) employed radioactive xenon scintillography to study middle ear ventilation, and reported preliminary evidence that this method distinguishes patients with otitis media.

Gas exchange has also been examined in the middle ear cleft. Elner (17) recently summarized this process, and pointed out the special characteristics of the middle ear as a gas pocket. Normal tubal function results in equalization of air entering and leaving the middle ear. Air leaves the middle ear by tissue absorption, and there is negligible gas exchange through the intact tympanic membrane. With blockage of the eustachian tube there is increasing negative pressure within the middle ear. At a certain unknown point, transudation of fluid occurs into the middle ear cavity, reducing the air volume of the cavity and relieving the pressure difference between cavity and surrounding tissues.

Paparella et al (18) investigated the effects of experimental eustachian tube obstruction in the squirrel monkey. Acutely there was widening of the subepithelial space due to accumulation of fluid. This was followed by rupture of the basement membrane, with fluid passing between epithelial cells into the middle ear cavity. Biochemical studies showed increased oxidative and hydrolytic enzyme activity throughout the mucoperiosteum. Physiologic studies provided direct evidence of vascular dilatation of the subepithelial small vessels in this experimental model (19). These investigators employed similar methods for tubal obstruction in the cat and chinchilla, and obtained results comparable to the monkey (20). These animal models have provided insights into the pathogenesis of effusion following tubal dysfunction in humans.

Microbial

The bacteriology of middle ear aspirates from children with acute otitis media have been the most widely reported. Little variance has been observed in the frequency of recovering pathogenic bacteria from these effusions over the past 25 years (Figure 1). In children age one year and older, *Streptococcus pneumoniae* has been cultured from 25–50% of effusions and *Hemophilus influenzae* from 15–25% (21–27). Of the 88 known pneumococcal serotypes, only seven types account for 70% of middle ear infection with the pneumococcus (28). These seven (types 1, 3, 6, 14, 18, 19, and 23) are also frequent isolates in patients with pneumococcal pneumonia, bacteremia, and meningitis (29–31). In contrast, *H. influenzae* isolates from middle ears are rarely (< 10%) type b, most being nontypeable, while hemophilus isolates in other instances of invasive disease are nearly always type b (32, 33). The incidence of pneumococcal and hemophilus otitis media remains relatively constant between 6 weeks and 13 years of age; hemophilus otitis does not decrease in frequency after age four years (27, 34). This is surprising in light of previous reports (21, 23, 25). While beta-streptococci accounted for a significant proportion of effusion isolates 25 years ago (21), this organism has been isolated from less than 6% of effusions in all recent reports. Other bacteria cultured from the effusions of children with acute otitis media include *Neisseria catarrhalis* in 2–9% of aspirates (24, 26, 27) and occasionally *S. epidermidis* (35). While most investigators accept the pathogenic role of *N. catarrhalis* in the middle ear, many have questioned the significance of *S. epidermidis* in otitis media since this organism normally colonizes the auditory canal (27). Feigin et al (35) however, were able to isolate *S. epidermidis* in pure culture from purulent middle ear effusions in 10 of 130 children with acute otitis media. All ten patients were febrile and showed clinical signs of middle ear inflammation. In addition, organisms were seen within polymorphonuclear leukocytes on gram stain in nine of the ten effusions.

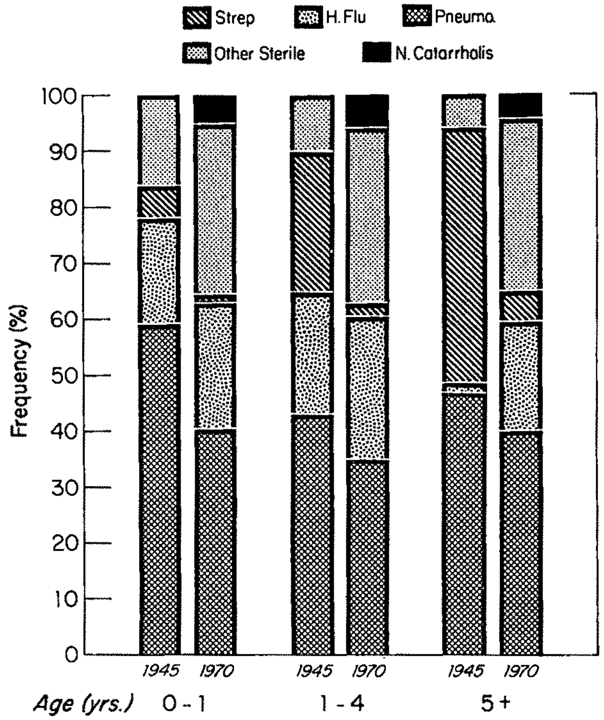


Figure 1 Bacteriology of acute otitis media compared by age group over a 25-year interval.

In contrast to these observations on the bacteriology of acute otitis in infants and children, the effusions of neonates are frequently colonized with coliform bacteria and *S. aureus* (36, 37). Common respiratory pathogens also play a significant role in acute otitis media even in young neonates. Tetzlaff et al (38) reported the recovery of pneumococci and hemophilus in 67% of effusions from 44 neonates, and of coliforms in 23% of these samples. Coliforms were not isolated from acute effusions in infants older than six weeks. These reports emphasize the importance of acute-otitis media during the neonatal period. McLellan et al (39) believe that otitis media, especially in the premature, may be a frequent cause of bacteremia, and that the tympanic membranes of sick neonates must be carefully examined.

Many investigators have obtained a culture of the nasopharynx coincident with the effusion sample, since the nasopharynx is a more accessible source of potential pathogens (23, 26, 40–44). Unfortunately, in 50–80% of these cases, more than one pathogen was recovered from the nasopharynx, which made correlation with either culture-positive or sterile effusion impossible. Although in most instances the organism present in the middle ear was also present in the nasopharynx, the multiplicity of potential nasopharyngeal pathogens does not permit upper respiratory cultures to accurately predict the middle ear pathogen.

Despite the evidence for a bacterial etiology in approximately 70% of episodes of acute otitis media, nearly one third of these effusions have been sterile when cultured by aerobic methods. Anaerobic bacteria have only recently been sought in these effusions. In a recent report, anaerobes were isolated in 35% of children with acute otitis media (45). However, organisms were rarely seen on gram stain. Studies of gas tension in middle ear effusion have shown an average Po_2 of 41 mm Hg, which suggests this is an unlikely site for anaerobic infection (46).

Attempts to prove a viral or mycoplasma etiology were recently reviewed by Klein & Teele (47). Viruses have been isolated from the effusions of 29 of 663 patients cultured (4.4%). A bacterial pathogen was found in 15 of these 29 fluids. However, nearly 24% of patients had nasopharyngeal viral colonization, and seroconversion to one or more virus occurred in 29% of patients. *Mycoplasma pneumoniae* has been recovered from middle ear effusion in only one of 771 patients cultured. Three children with effusions had mycoplasma isolated from the throat or nasopharynx. Mycoplasma seroconversion, however, occurred in 9% of patients. Virus and/or mycoplasma may only infect the nasopharynx, which results in eustachian tube blockage and transudation of fluid into the middle ear. Alternatively, these microbial agents may be associated with infection during an early stage but disappear by the time of acute symptomatology, may be difficult to isolate due to an inhibitory material in the effusion, or may be present but methods for isolation have not been adequate.

The search for a nonbacterial etiology for acute otitis media has included culture of effusions for bacterial L forms. Since lysozyme, which induces L forms in vitro, is found in these effusions, it is reasonable to assume that cell-wall defective bacteria might be present. Kamme (40) was able to isolate *S. pneumoniae* only on L-form medium from 2 of 17 "sterile" middle ear effusions. Thus, it is conceivable that reversion to ordinary bacterial forms may occasionally occur with resulting clinical relapse.

Chronic, nonsuppurative forms of otitis media have only recently been examined for bacteria. Liu et al (48) examined 102 chronic effusion samples by culture and gram- and Wright stain: specimens were classified as serous, mucoid, or leukocytic by physical and cytologic characteristics. Leukocytic effusions were defined as those chronic effusions that contained polymorphonuclear leukocytes. Seventy-five percent of effusions showed bacteria on smear, and 52% yielded bacteria on culture. Younger patients had positive cultures more frequently than older patients, which suggests that these effusions may become sterile with increasing time. Mucoid effusions had 37% positive cultures, while 59% of serous effusions and 64% of leukocytic effusions cultured bacteria. Eleven different bacterial types were cultured. Common pathogens (*H. influenzae*, *S. pneumoniae*, group A streptococci, *S. aureus* and *epidermidis*, and *N. catarrhalis*) comprised the majority of positive cultures. Nearly all positive cultures yielded a single bacterial species. In serous effusions, *S. epidermidis* was most frequently isolated, and *H. influenzae* was most common in leukocytic effusions. Fifty percent of the culture-positive mucoid effusions yielded either pneumococcus or hemophilus.

We have made similar observations in 68 children with chronic otitis media (49). Effusion cultures were positive in 47% of children with effusion. *H. influenzae* was most frequently isolated in children with serous effusion. In contrast, *S. epidermidis* was isolated from mucoid effusions. Most serous effusions contained polymorphonuclear leukocytes, while mucoid effusions contained fibrinous eosinophilic strands and desquamated epithelial cells, but were devoid of phagocytic cells. These data provide evidence that chronic bacterial infection may be an important aspect of the pathogenesis of chronic serous and mucoid otitis media.

Immunochemical

The pathogenesis of otitis media has recently been approached from both its immunologic and biochemical aspects. The middle ear is the site of a distinct, secretory immune system similar to that found in other areas of the upper respiratory tract. Initial reports have described the presence of immunoglobulin in middle ear effusion in higher concentrations than in simultaneous serum samples (28, 50). Appreciable amounts of IgA (7S and 11S) have been found in middle ear effusion. An indirect relationship exists between high concentrations of effusion IgA and recovery of pathogenic bacteria from the middle ear (48–53). We and others have observed that the concentration of IgA in mucoid effusions is significantly higher than in serous effusions (48, 49). Fluorescent antibody staining of middle ear mucosa has demonstrated secretory component in the epithelial layer (51, 53), and radioimmunoelectrophoresis has detected free secretory component in as many as 20% of effusions (52). Plasma cells, seen in the subepithelial mucosal interstitium, have been coated primarily with IgA (51). In addition, middle ear mucosa grown in vitro synthesized IgA, IgG, and secretory component (53).

The functional significance of this local secretory defense system has been demonstrated by showing that patients with acute otitis media due to *S. pneumoniae* or *H. influenzae* who develop specific antibody in their middle ear effusion clear bacteria (28, 54, 55). All three classes of antibody (IgG, IgA, and IgM) have been found in these effusions, and the clearing of bacteria does not appear to be associated with the immunoglobulin class. Thus, an etiopathogenic relationship appears to exist between the invading bacteria and local and/or systemic immune defenses.

Allergy has been implicated in the pathogenesis of otitis media. However, most investigators have been unable to demonstrate increased IgE concentrations in serum or in effusion (48, 56–60). In our own inventory of 68 children with chronic otitis media, over half of the children had symptoms suggestive of clinical upper respiratory allergy (49).

The biologic effects of certain bacterial products have also been considered to play a role in inflammation and host immune responses. Coonrod & Drennan (61) have shown that when the serum from patients with pneumococcal sepsis contained capsular polysaccharide antigen, development of type-specific antibody was delayed and there was protracted illness. Several laboratories, including our own (G. S. Giebink, unpublished observations), have demonstrated that pneumococcal polysaccharide activates the alternative complement pathway in vitro and in vivo, reduces circulating C3, and results in a potential defect in pneumococcal opsonization, phagocytosis, and microbial killing (62–64).

Middle ear effusions also contain other mediators of inflammation, including chemotactic factor, macrophage inhibition factor, activated-complement components, prostaglandins, and oxidative and hydrolytic enzymes (48, 65–68). The significance of these factors in the pathogenesis of middle ear disease is uncertain. However, tissue destruction and bone resorption, characteristic of chronic otitis media, are probably related to persistence of these active factors (69).

Recent reports suggest that patients with chronic otitis media may have subtle underlying abnormalities in systemic immunity. Johnson et al (70) found unusual complexes of C1 subcomponents (C1r and C1s) in sera of 9 of 14 children with chronic otitis and in only 1 of 6 children with acute otitis. We have observed depressed peripheral neutrophil chemotaxis in 11 of 68 children (16%) with chronic otitis media (49). Hill et al (71) described 14 children with chronic otitis media and chronic diarrhea who had a marked reduction in their peripheral neutrophil chemotactic activity. Nine children with only chronic diarrhea and ten children who had only otitis showed normal neutrophil chemotaxis. Defective chemotaxis could not be detected in any children after the effusions cleared, which suggests that depressed chemotaxis is a secondary, transient phenomenon. Normal serum immunoglobulin levels in patients with otitis media have been documented; normal systemic cell-mediated immune responses and normal peripheral neutrophil phagocytic and bactericidal capacity have been described (49, 71).

Because of the multitude of variables encountered in studying human immunologic and biochemical responses to otitis media, animal models have been developed that permit detailed investigation. Paparella et al (18) introduced a method for producing serous effusions in monkeys. Hussl & Lim (72) employing this model, found increased numbers of IgG-, IgA-, IgM-, and IgE-coated plasma cells in middle ear mucosa in experimental serous otitis (72). The cat and, most recently, the chinchilla have also been employed as animal models for otitis media (20). Following eustachian tube obstruction, both species show early production of a serous effusion with considerable epithelial and subepithlial activity, shown both histologically and biochemically. As the effusion persists, it becomes mucoid in quality, and both total protein and hexosamine content increase, which reflects the secretory nature of mucoid otitis.

We have developed an animal model with chinchillas for investigating acute pneumococcal otitis media, and have measured the immunopathologic responses to pneumococcal serotypes 3, 6, 7, 18, and 23 in this model (73, 74). Animals developed purulent middle ear effusions following direct inoculation of small numbers of bacteria (as few as ten bacteria per ear) into their middle ears. Without treatment, the infection persisted for three to four weeks and was followed by clearing of bacteria. In many instances granulation tissue persisted within the ear for up to six weeks after infection. Type-specific serum pneumococcal antibody increased within two weeks following infection; however, in most animals, effusion antibody concentration remained equal to or less than the serum concentration. The concentration of effusion antibody exceeded serum in occasional animals, which suggests a local immune response. Since the pneumococcal polysaccharide antigen content of these effusions is high, antigen-antibody complexing may affect the ability to detect free antibody in most animals. Biochemical assay of these experimental effusions revealed high levels of several oxidative and hydrolytic enzymes after infection. With clearing of bacteria from the middle ear, these enzymes decreased in concentration but lysozyme persists, which suggests a continuing inflammatory process. We are utilizing this animal model to evaluate the effect of pneumococcal polysaccharide vaccine in preventing the experimental middle ear infection.

DIAGNOSIS

Clinical Symptoms and Signs

Presenting clinical symptoms in acute otitis media are variable, nonspecific, and, particularly in infants, often misleading (24, 27, 75–77). While upper respiratory symptoms are usually present, gastrointestinal complaints or irritability may be the only symptoms of severe otitis media in younger children and infants. Indeed, 50% have complaints referrable to the ear (e.g. otalgia, ear tugging, otorrhea). Howie et al (27) recorded clinical data in 858 instances of acute otitis media, and noted that pneumococcal otitis media was significantly more likely to be associated with severe otalgia and high fever then was hemophilus otitis. Hemophilus otitis tended to be associated with little fever or pain, was more apt to be bilateral, and was significantly associated with symptoms of rhinitis for two or more weeks prior to diagnosis (27). Chronic otitis media is often completely asymptomatic and only detected by otoscopy or routine audiometric screening. A recent report suggests that 20% of preschool children may have painless middle ear effusions (78).

Otoscopy and Tympanometry

Since most cases of acute otitis media occur during infancy and early childhood, at a time when localization of symptoms is most difficult, careful examination of the tympanic membranes is required. Most clinicians and researchers accept that a bulging, diffusely red tympanic membrane is an accurate predictor of middle ear infection. Halstad et al (25) confirmed this impression by culturing pathogens from the middle ears of 61 of 67 patients with bulging red eardrums. In contrast, none of 15 patients whose eardrums were diffusely red but were otherwise normal in appearance contained pathogens. Impaired mobility of the tympanic membrane is another diagnostic sign indicating middle ear effusion (79).

Tympanometry has been proposed as a means for objectively examining the tympanic membrane (80–83). Tympanometry measures the absorption of sound (acoustic impedance), which is least when the middle ear and external canal air pressures are equal. During testing, the external canal pressure is artificially increased and decreased, and sound absorption is recorded as a tympanogram. Tympanic membrane scarring or middle ear effusion increase impedance. Tympanometry has been shown to be 93% accurate in detecting effusion in 70 children age 2–15 years (80). In a larger study of 280 children age ten days to five years, tympanometry successfully detected middle ear effusion in 86% of children over age seven months (83). In infants younger than seven months, the tympanogram is frequently normal despite the presence of effusions. Newer methodology, however, promises improved results in young infants (82). Audiometry is useful for measuring sensory-neural hearing loss, but is not helpful in detecting middle ear effusion. Normal audiograms are recorded in as many as 30–50% of children who have abnormal tympanograms and effusion at myringotomy (84–86). Tympanometry may supplement otoscopy in the diagnosis of otitis media and may by particularly helpful in evaluating therapy.

Tympanocentesis

Tympanocentesis is often necessary to confirm the presence of effusion in suspected cases of otitis media or to culture for bacterial pathogens. Pathogens are frequently recovered from chronic effusions; therefore this procedure is useful in chronic as well as acute cases. Indications for tympanocentesis include: (*a*) otitis media in neonates where it is important to establish a potential etiologic agent for sepsis; (*b*) otitis media in immunologically compromised patients to permit narrow-spectrum antibiotic treatment; (*c*) otitis media in seriously ill patients where suppurative complications, such as mastoiditis, meningitis, or brain abscess, are suspected; and (*d*) unsatisfactory response to antimicrobial therapy with persisting otitis media. Treatment failure may be due to ampicillin-resistant *H. influenzae* as these organisms have been recovered from increasing numbers of children with otitis media (87).

MANAGEMENT

Antimicrobial

Antibiotic treatment has dramatically reduced the morbidity of acute, purulent otitis media (88–91). During the preantibiotic era, Bakwin & Jacobziner (88) and Heller (89) reported that the incidence of mastoiditis complicating purulent otitis was 2–6%. When effective antibiotics became available Heller found no cases of mastoiditis among 1036 children with purulent otitis (90). The effect of antibiotics on acute otitis has been most clearly demonstrated in studies employing placebo-treated groups (25, 91, 92). Placebo-treated children frequently have resolution of symptoms, but otorrhea persists in the majority. Placebo-treated patients have shown persistence of pneumococci in 36 of 45 patients (80%) and persistence of hemophilus in 12 of 21 (57%) (91). A great difference was found between antibiotics and placebo in clearing pneumococcus and hemophilus from the middle ear.

Selection of appropriate antimicrobial agents for acute otitis media requires knowledge of the pathogen and its antimicrobial sensitivity pattern, and the ability to concentrate the agent in the middle ear effusion. Antimicrobial drug trials therefore need to include bacterial culture of middle ear effusion. Studies that have employed bacteriology show generally good agreement between the efficacy of various antimicrobial regimens and in vitro sensitivity of the infecting organisms to commonly used antibiotics (25, 26, 91, 93, 94). In clinical practice, however, it would be impractical to perform tympanocentesis in all patients with otitis media. Since *S. pneumoniae* and *H. influenzae* comprise over 50% of pathogens isolated from acute effusions, antimicrobial regimens selected must cover these organisms when specific bacterial etiology is unknown.

Table 2 illustrates the minimal inhibitory concentration (MIC) for four antimicrobial agents required to inhibit 90% of *S. pneumoniae* and *H. influenzae* (94). Of 14 hemophilus tested for sulfonamide sensitivity, 11 (79%) were sensitive. Pneumococcal susceptibility to sulfonamide was not tested. Ear effusion was obtained from 30 minutes to two hours after oral antibiotic administration to determine antibiotic concentration (Table 2). Comparison of MIC with mean effusion antibiotic concentration illustrates that 90% of the pneumococci would be inhibited in the middle ear by penicillin, amoxicillin, or erythromycin. However, less than 25% of hemophilus in the ear would be inhibited by either penicillin or erythromycin; sulfonamide alone would have inhibited all eleven sensitive strains. Penicillin entered the effusions least effectively, and sulfonamide had the highest effusion to serum ratio. Coffey (95) found a somewhat higher ratio of effusion to serum (0.73) among ten patients with acute otitis one hour following an intramuscular injection of ampicillin. Others, however, have shown that the effusion antibiotic concentration may be considerably lower in patients with mucoid otitis (96).

Table 2 Antibiotic susceptibilities and middle ear effusion concentration[a]

Drug	Minimum inhibitory concentration (μg/ml)		Mean antibiotic concentration in effusion (range)	Ratio of effusion to plasma
	Pneumococcus	*Hemophilus*		
Penicillin	0.04	0.63	0.2 μg/ml (<0.03–0.3)	0.13
Amoxicillin	0.01	0.63	2.2 μg/ml (<0.03–4.1)	0.46
Erythromycin estolate	0.08	2.5	1.3 μg/ml (<0.1–3.2)	0.33
Sulfonamide	(not tested)	[b]	8.3 mg/dl (4.4–12)	0.62

[a] Adapted from Howard et al. See Ref. 94.

[b] Eleven of 14 hemophilus strains were sensitive to sulfonamide (MIC = 0.1–2.5 mg/dl).

While only a few investigators have reaspirated ears to document the effectiveness of treatment (93), several researchers have performed careful, sequential otoscopic examinations (26, 91, 94, 97). The clinical success of the various antimicrobial regimes studied with reference to pneumococcal and hemophilus bacteriology is shown in Table 3. For pneumococcal otitis, only sulfonamide treatment was clearly less effective than any other single drug or drug combination. While most regimens were adequate for pneumococcal otitis, only ampicillin, penicillin V and sulfonamide, or erythromycin and sulfonamide cured more than 60% of patients with hemophilus otitis. Thus, when the pathogen causing otitis is not known, these data suggest that optimum treatment would include either ampicillin (or amoxicillin), erythromycin and sulfonamide, or penicillin V and sulfonamide. With the emergence of ampicillin-resistant *H. influenzae* (87), combination therapy may have advantages over ampicillin or amoxicillin alone. The usual duration of antimicrobial therapy is 10–14 days, although clinical trials have not been performed. No patient, however, should be considered cured without complete resolution of the signs of middle ear disease.

Antimicrobial therapy may also effect the immune response in otitis media. Preliminary findings suggest that a greater percentage of patients develop specific IgM antibody following treatment with erythromycin and sulfonamide (31%) than after penicillin (16%), ampicillin (18%), or penicillin and sulfonamide (24%) treatment (98). Whether these immune responses protect the individual from recurrent acute or chronic otitis media remains unknown at present.

Decongestant

Both medical and surgical regimens have been recommended to promote resolution of persisting middle ear effusions following acute otitis media and chronic otitis with effusion. Decongestants, both alone and in combination with antihistamines, are often used, but evidence for their effectiveness is scant. Rubenstein et al (76) were unable to show that pseudoephedrine altered the outcome of acute otitis; the same group showed that patients treated with chlorpheniramine

Table 3 Clinical success of various antimicrobial regimens for acute otitis media[a]

	No. Successes/total no. treated (% success)	
Drug	*S. pneumoniae*	*H. influenzae*
Penicillin V	22/30 (73)	6/25 (24)
Benzathine pen + procaine pen	8/9 (89)	0/7 (0)
Ampicillin	64/70 (91)	49/58 (84)
Erythromycin	20/22 (91)	5/28 (18)
Cephalexin	9/12 (75)	6/10 (60)
Sulfonamide	20/31 (65)	6/10 (60)
Pen V + sulfa	43/48 (90)	19/27 (70)
Erythro + sulfa	30/33 (91)	25/28 (89)

[a] Adapted from Refs. 26, 91, 93, 94, 97.

maleate did slightly better than those who received antibiotics alone (77). Unfortunately, bacteriologic investigation was not performed in these two large clinical studies. Thus, one is uncertain whether control or treatment groups were biased and contained a preponderance of infected or sterile effusions. Employing a positive- and negative-pressure test system to evaluate eustachian tube function, Miller (99) demonstrated more rapid equalization of middle ear pressures in 5 of 13 children with chronic otitis treated with a combination of pseudoephedrine and carbinoxamine maleate than with placebo. This combination might be especially effective in children with allergic rhinitis and chronic otitis.

Myringotomy

The value of surgical drainage of the middle ear in addition to antimicrobial treatment of acute otitis media has not been established. In comparing his experience prior to and following the introduction of antibiotics, Heller (89, 90) found a tenfold decrease in the need to perform myringotomy associated with a 34% increase in the number of patients with an uneventful recovery. Subsequently, Roddey et al (100) performed a controlled study on the possible benefit of myringotomy. All 181 children with acute otitis media were treated with benzathine penicillin, oral sulfonamide, and nasal phenylephrine; in addition, myringotomy was performed on 113 infected ears (86 patients) and not on 127 ears (95 patients). The only significant difference between the two groups, judged by otoscopy at 2, 10, 30, and 60 days, and by audiometry at 3–6 months, was more rapid pain relief among a small group who had severe otalgia initially. Studies have not been performed to evaluate the effect of myringotomy on immunologic or biochemical events in the middle ear. At the present time the chief indication for myringotomy at the onset of acute otitis media is persistent, severe otalgia not relieved by analgesics and antibiotics. Suppurative complications, such as mastoiditis, also require myringotomy for adequate drainage.

Tympanostomy Tubes

The era of middle ear ventilation via surgical placement of tympanostomy tubes began only 20 years ago. The purpose of the tube is to simulate eustachian tube function, which presumably is deficient in chronic or recurrent acute otitis media. Presently, this procedure is performed at a rate as high as 3 per 1000 population (101).

Recently, Kokko & Palva (102) reported the results of a controlled trial of tympanostomy tubes in Finland. Tubes were inserted in 164 patients (290 ears) and were not placed in 17 patients (33 ears). Indications for surgery included failure of antibiotic and myringotomy treatment. Average duration of effusion prior to tube placement was 8½ months, and tubes remained in place an average of 7 months. Mean age was 5½ years, although 31 children were under one year of age. Pathogens were found in the effusions in 22% of the cases.

At reexamination, an average of three years after surgery, 65% of the ears treated with tubes had healed completely, in 32% the appearance of the ear drum was abnormal at follow-up, and in 6.3% eardrum changes were associated with depressed hearing. Mean duration of medical management in the control group was six months. At reexamination, after an average of 4½ years of medical management, 26% of these eardrums were abnormal; in 6.9% the changes were associated with impaired hearing.

Complications of tympanostomy tube treatment have occurred in 21–27% of cases depending on the period of observation after insertion (102, 103). Persistent eardrum perforation with or without ear drainage may occur in as many as 15% of patients treated, and occasionally leads to otitis externa (103). Middle ear granulation was noted in 0.1% of treated ears in one series (103), and cholesteatoma in 1.4% of ears in another report (102).

These reports cast some doubt on the value of tympanostomy tube treatment in altering the long-term outcome of chronic and recurrent acute otitis media. However, it will be necessary to perform larger controlled studies matching age, bacteriology, and both prior and subsequent medical treatment. Language development, as well as tympanometry, audiometry, and early and late complications must be assessed. Until that time, precise indications for tube treatment cannot be made.

Tonsillectomy and Adenoidectomy

Tonsillectomy and adenoidectomy, or adenoidectomy alone, is performed with great frequency in patients with otitis media. Indeed, in 1973 these procedures were performed at a rate of 150–200 per 10,000 population (104). While complications are infrequent, the overall mortality from these operations is 1 per 16,000–27,000 operations, and nonfatal hemorrhage occurs in approximately 1 of 3500 procedures (105).

Tonsillectomy alone does not appear to alter the course of otitis media (106). The efficacy of adenoidectomy in managing children with otitis media has not been demonstrated although a prospective, carefully designed, controlled study is currently in progress (107). Previous trials have yielded conflicting results. McKee (106) and Roydhouse (108) found a reduced incidence of otitis after adenoidectomy, while Brownlee et al (7) and Mawson et al (109) did not. These studies were impaired by a failure to separate adenoidectomy from tonsillectomy and adenoidectomy or from tonsillectomy alone, lack of stringent criteria for surgery, inadequate recognition of otitis media, and poor methods for follow-up evaluation.

At the present time indications for adenoidectomy in children who have only chronic or recurrent acute otitis media without upper airway obstructive disease are ill-defined. Serial audiometric, and possibly tympanometric, testing prior to surgery seems advisable. Failure of adequate medical therapy may warrant the procedure. Contraindications include velopharyngeal insufficiency, blood dyscrasia or coagulopathy, and probably allergic rhinitis.

Prophylaxis

Otitis media is one of the most common of the severe diseases of childhood, and is frequently a recurrent disease, especially among infants and preschool children (8). Antibiotic prophylaxis has been shown to be effective in reducing subsequent episodes of acute otitis media (110–112). Sulfamethoxypyridazine, a long-acting sulfonamide, was shown to reduce the frequency of otitis media in an American Indian population (110). Recently, these observations were confirmed in a double-blind crossover study (111). The attack rate in these patients while on placebo equalled the number of episodes per child during the year prior to study. Sulfisoxazole was shown to significantly reduce the frequency of otitis among children less than five years of age. Sulfisoxazole dosage was 500 mg twice daily for all children, age 11 months to eight years. Ampicillin prophylaxis may also be effective, since a single daily dose (125 mg for children up to 2½ years; 250 mg for those older) has been shown to reduce the incidence of otorrhea by 47% in a double-blind study of 364 Alaskan Eskimo children (112). Among children where ampicillin compliance was high, the reduction in incidence was 67%.

The propensity for recurrent acute otitis media appears to be established during the first year of life, and is related to an initial episode of pneumococcal otitis media (8). Since only a few pneumococcal serotypes are associated with this disease, another means to reduce the number of patients susceptible to recurrent otitis may be afforded by a polyvalent pneumococcal vaccine. A preliminary study by Howie et al (113) showed that among 60 children given an octavalent pneumococcal vaccine, only 11 (18%) subsequently developed vaccine type otitis media compared to 6 of 13 (46%) given saline placebo. Age at vaccination appeared to affect protection. Only 1 of 26 (4%) children vaccinated between ages 6 and 19 months developed vaccine type otitis, while the frequency of vaccine type otitis media in a control group of similar age was 30% (19 of 64 patients). Only minimal local side effects are occasionally encountered with the vaccine. Presently, pneumococcal vaccine field trials are being conducted in several centers to determine its value in preventing recurrent acute otitis media.

COMPLICATIONS AND SEQUELAE

Following diagnosis and initiation of antimicrobial therapy, the majority (60–90%) of patients are improved within 72 hours. It is unusual for patients to remain symptomatic at 10–14 days, although as many as 30% may have abnormal eardrums and persistent middle ear effusions (25, 76, 78, 93, 94, 100). In approximately half of these cases, the same organism will be cultured at follow-up that was isolated initially, which suggests treatment failure or poor compliance (25, 93, 94). Repeated attacks of acute otitis are quite common, especially with those children who have their initial episode in the first one to two years of life (8, 44). *S. pneumoniae* are most often isolated from the effusions of these children during the initial bout (8). Repeated episodes, however, may be caused by any of the common pathogens (24, 25, 94). Suppurative complications of acute otitis media, such as mastoiditis, meningitis, lateral sinus thrombosis, brain abscess, and subdural empyema, are rare events today except in neglected cases (76, 77, 88–90, 114–116). Children with pyogenic meningitis, however, often have otitis media (114).

The natural history of chronic otitis media with persisting middle ear effusions is unknown. Short-term studies have shown that the presence of effusion is associated with hearing loss (77, 100, 117). Routine audiometry suggests that adequately treated acute otitis media does not lead to permanent hearing loss (7). Long-term studies performed on children with less than adequate treatment indicate that 15–55% may have permanent hearing loss (2, 118, 119). It is unknown whether these children had single episodes of acute otitis, recurrent acute otitis, or chronic otitis. It has been recently suggested that acute otitis may result in permanent ultra-high-frequency hearing loss, and that these effects might be additive with subsequent infections (120).

A few studies suggest that mild to moderate hearing impairments, even if relieved subsequently, have lasting effects on speech and language development (2, 121, 122). Presumably, the prior episodes of otitis media, which occur during formative years, distort auditory processing and integration of these learning stimuli. However, these studies come from highly selected populations, and data on prior middle ear disease was collected retrospectively. It is not known whether these children represent the trend for all children with otitis media or whether they represent an extreme situation. Only comprehensive, long-term prospective studies will answer these questions.

CONCLUSIONS

Clinical and experimental studies of both suppurative and nonsuppurative types of otitis media suggest that this disease is multifaceted, and that clinical and pathologic changes fall along a continuum. Epidemiologic data provide evidence that the majority of children in developed as well as developing countries have at least one bout of acute, suppurative otitis media during the first two years of life. A great many of these children, especially those who have an episode of pneumococcal otitis media prior to their first birthday, suffer recurrent otitis media. Although the frequency of acute otitis diminishes with age, several point-prevalence studies of preschool- and school-aged children show that middle ear effusions, i.e. nonsuppurative, serous, mucoid, and chronic otitis media, may be present with little clinical symptomatology. The percentage of children with acute otitis media during their early years that develop nonsuppurative otitis is not known. However, recent studies of linguistic skills suggest that some children with prior middle ear disease develop lasting speech and language defects.

The pathogenesis of otitis media is being explored in several laboratories with clinical studies and with experimental animals. The recent discovery of pathogenic bacteria in persistent serous and mucoid middle ear effusions suggests that microbial factors may be related to the pathogenesis of chronic, as well as acute, otitis media. Several biochemical mediators of inflammation are found in these effusions, and their concentrations tend to increase with persistence of effusion and with the development of mucoid effusion. In addition, the middle ear is the site of a distinct secretory immune system. Eustachian tubal function is also directly related to the pathogenesis of otitis, yet it is not known whether tubal dysfunction is a primary or secondary event. The composite of these observations suggests that certain patients are unable to coordinate effective mechanical, biochemical, and immunologic defenses in the face of middle ear inflammation. Whether the defect in host response is induced by a particular microbial pathogen or is inherent in the host remains a mystery. Purulent otitis media, which progresses to chronic otitis, is produced in chinchillas by direct inoculation of *S. pneumoniae* and serous otitis can be studied in cats and other animals. Sequential analysis of single variables in these experimental animals should provide answers to the pathogenesis of the common disease.

Until more is known of pathogenesis, and methods are devised that control the disease process early in its development, antimicrobial agents must continue to be the mainstay of therapy for patients with acute otitis media. Antimicrobial agents such as sulfonamides also appear to be beneficial in prophylaxis of acute otitis. Selection of appropriate antimicrobials is based on patient age, clinical signs and symptoms, knowledge of the microbial flora in effusions, penetration of these agents into effusions, and, possibly, on the relationship of certain therapeutic agents to local immune responses.

Protection from otitis media may also become available from immunologic procedures. Recent clinical testing of an octavalent pneumococcal vaccine in children older than six months suggests that this may provide effective prophylaxis for otitis media.

With the recognition of the otitis medias as a serious cause of childhood morbidity, there is a renewed interest in this disease and increased research efforts in several medical centers. This concern and the combined efforts of clinicians and basic scientists has led to greatly increased knowledge of pathogenesis. This knowledge should lead to more effective methods for diagnosis and management.

Literature Cited

1. Paparella, M. M. 1976. Summation of symposium. Recent advances in middle ear effusions. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 296–299
2. Kaplan, G. J., Fleshman, J. K., Bender, T. R., Baum, C., Clark, P. S. 1973. Long-term effects of otitis media. A ten-year cohort study of Alaskan Eskimo children. *Pediatrics* 52:577–85
3. Tonkin, S. 1970. Maori infant health: II. Study of morbidity and medico-social aspects. *N. Z. Med. J.* 72:229–38
4. Jose, D. G., Self, M. H. R., Stallman, N. D. 1969. A survey of children and adolescents on Queensland Aboriginal settlements. *Aust. Paediatr. J.* 5:71–88
5. Mortimer, E. A. 1973. Indian health: An unmet problem. *Pediatrics* 51:1065–66
6. Report of a survey by the Medical Research Council's Working-Party for Research in General Practice. 1957. Acute otitis media in general practice. *Lancet* 2:510–14
7. Brownlee, R. C., DeLoache, W. R., Cowan, C. C. Jr., Jackson, N. P. 1969. Otitis media in children: Incidence, treatment, and prognosis in pediatric practice. *J. Pediatr.* 75:636–42

8. Howie, V. M., Ploussard, J. H., Sloyer, J. L. 1975. The "otitis-prone" condition. *Am. J. Dis. Child.* 129:676–78
9. Sklonik, E. M. 1958. Otologic evaluation in cleft palate patients. *Laryngoscope* 68:1908–49
10. Paradise, J. L., Bluestone, C. D., Felder, H. 1969. The universality of otitis media in 50 infants with cleft palate. *Pediatrics* 44:35–42
11. Paradise, J. L., Bluestone, C. D. 1974. Early treatment of the universal otitis media of infants with cleft palate. *Pediatrics* 53:48–54
12. Pelton, S. I., Shurin, P. A., Klein, J. O. 1977. Persistence of middle ear effusion after otitis media. *Pediatr. Res.* 11:793 (Abstr.)
13. Cantekin, E. I., Bluestone, C. D., Parkin, L. P. 1976. Eustachian tube ventilatory function in children. *Ann. Otol. Rhinol. Laryngol.* 85: Suppl. 25, pp. 171–77
14. Holmquist, J., Renvall, U. 1976. Middle ear ventilation in secretory otitis media. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 178–81
15. Ingelstedt, S. 1976. Physiology of the eustachian tube. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 156–60
16. Kirchner, F. R., Robinson, R., Smith, R. F. 1976. Study of the ventilation of middle ear using radioactive xenon. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 165–68
17. Elner, Å. 1976. Normal gas exchange in the human middle ear. *Ann. Otol. Rhinol. Laryngol.* 85: Suppl. 25, pp. 161–64
18. Paparella, M. M., Hiraide, F., Juhn, S. K., Kaneko, Y. 1970. Cellular events involved in middle ear fluid production. *Ann. Otol. Rhinol. Laryngol.* 79:766–79
19. Hiraide, F., Paparella, M. M. 1972. Vascular changes in middle ear effusions. *Arch. Otolaryngol.* 96:45–51
20. Juhn, S. K., Paparella, M. M., Kim, C. S., Goycoolea, M. V., Giebink, S. 1977. Pathogenesis of otitis media. *Ann. Otol. Rhinol. Laryngol.* 86:481–92
21. Nielsen, J. C. 1945. *Studies of the aetiology of acute otitis media,* Copenhagen: Munksgaard
22. Grönroos, J. A., Vihma, L., Salmivalli, A., Berglund, B. 1968. Co-existing viral (respiratory syncytial) and bacterial (pneumococcus) otitis media in children. *Acta Otolaryngol* 65:505–17
23. Feingold, M., Klein, J. O., Haslam, G. E. Jr., Tilles, J. G., Finland, M., Gellis, S. M. 1966. Acute otitis media in children. Bacteriologic findings in middle ear fluid obtained by needle aspiration. *Am. J. Dis. Child.* 111:361–65
24. Coffey, J. D. Jr., Booth, H. N., Martin, A. D. 1966. Otitis media in the practice of pediatrics. *Pediatrics* 38:25–32
25. Halsted, A., LePow, M. L., Balassanian, N., Emmerich, J., Wolinsky, E. 1968. Otitis media: Clinical observations, microbiology, and evaluation of therapy. *Am. J. Dis. Child.* 115:542–51
26. Nilson, B. W., Poland, R. L., Thompson, R. S., Morehead, D., Baghdassarian, A., Carver, D. H. 1969. Acute otitis media: Treatment results in relation to bacterial etiology. *Pediatrics* 43:351–58
27. Howie, V. M., Ploussard, J. H., Lester, R. L. 1970. Otitis media: A clinical and bacteriological correlation. *Pediatrics* 45:29–35
28. Sloyer, J. L., Howie, V. M., Ploussard, J. H., Amman, A. J., Austrian, R., Johnston, R. B. Jr. 1974. Immune response to acute otitis media in children. I. Serotypes isolated and serum and middle ear fluid antibody in pneumococcal otitis media. *Infect. Immun.* 9:1028–32
29. Austrian, R., Gold, J. 1964. Pneumococcal bacteremia with especial reference to bacteremic pneumococcal pneumonia. *Ann. Intern. Med.* 60:759–76
30. Lund, E. 1970. Types of pneumococci found in blood, spinal fluid and pleural exudate during a period of 15 years (1954–1969). *Acta Pathol. Microbiol. Scand.* 78(B): 333–36
31. Finland, M., Barnes, M. W. 1977. Changes in occurence of capsular serotypes of *Streptococcus pneumoniae* at Boston City Hospital during selected years between 1935 and 1974. *J. Clin. Microbiol.* 5:154–66
32. Harding, A. L., Anderson, P., Howie, V. M., Ploussard, J. H., Smith, D. H. 1973. In *Hemophilus Influenzae,* ed. S. H. W. Sell, D. T. Karzon, pp. 21–28. Nashville, Tennessee: Vanderbilt Univ. Press. 325 pp.
33. Sell, S. H., Turner, D. J., Feder-

spiel, C. F. 1973. See Ref. 32, pp. 3–12

34. Gröroos, J. A., Kortekangas, A. E., Ojala, L., Vuori, M. 1964. The aetiology of acute middle ear infection. *Acta. Otolaryngol.* 58: 149–158

35. Feigin, R. D., Shackelford, P. G., Campbell, J., Lyles, T. O., Schechter, M., Lins, R. D. 1973. Assessment of the role of *Staphylococcus epidermidis* as a cause of otitis media. *Pediatrics* 52:569–76

36. Warren, W. S., Stool, S. E. 1971. Otitis media in low–birth–weight infants. *J. Pediatr.* 79:740–43

37. Bland, R. D. 1972. Otitis media in the first six weeks of life: Diagnosis, bacteriology, and management. *Pediatrics* 49:187–97

38. Tetzlaff, T. R., Ashworth, C., Nelson, J. D. 1977. Otitis media in children less than 12 weeks of age. *Pediatrics* 59:827–32

39. McLellan, M. S., Strong, J. P., Abdo, C. J., Tomberlin, C. G. 1968. Otitis media in a premature infant. *Am. J. Dis. Child.* 116:439–41

40. Kamme, C., Lundgren, K. Mårdh, P. 1971. The aetiology of acute otitis media in children. Occurrence of bacteria, L forms of bacteria and mycoplasma in the middle ear exudate. Relationship between bacterial findings in the middle ear exudate, nasopharynx and throat. *Scand. J. Infect. Dis.* 3:217–23

41. Jones, M. 1960. Comparison of bacteria from ear and upper respiratory tract in otitis media. *Arch. Otolaryngol.* 72:329–32

42. Mortimer, E. A. Jr., Watterson, R. L. Jr. 1956. A bacteriologic investigation of otitis media in infancy. *Pediatrics* 17:359–66

43. Howie, V. M., Ploussard, J. H. 1971. Simultaneous nasopharyngeal and middle ear exudate cultures in otitis media. *Pediatr. Dig.*, Feb 1971, pp.31–35

44. Branefors–Helander, P., Dahlberg, T., Nylén, O. 1975. Acute otitis media: A clinical, bacteriological and serological study of children with frequent episodes of acute otitis media. *Acta Otolaryngol.* 80:399–409

45. Brook, I., Anthony, B., Finegold, E. M. 1976. *Aerobic and anaerobic bacteriology of acute otitis media in children.* Presented at Intersci. Conf. Antimicrobial Agents Chemother., 16th, Chicago

46. Ingelstedt, S., Jonson, B., Rundcrantz, H. 1975. Gas tension and pH in middle ear effusion. *Ann. Otol. Rhinol. Laryngol.* 84:198–202

47. Klein, J. O., Teele, D. W. 1976. Isolation of viruses and mycoplasmas from middle ear effusions: A review. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 140–44

48. Liu, Y. S., Lim, D. J., Lang, R. W., Birck, H. G. 1975. Chronic middle ear effusions. *Arch. Otolaryngol.* 101:278–86

49. Giebink, G. S., Mills, E. L., Huff, J. S., Juhn, S. K., Quie, P. G. 1976. Host responses in otitis prone children. *Pediatr. Res.* 10:513 (Abstr.)

50. Howie, V. M., Ploussard, J. H., Sloyer, J. L. Jr., Johnston, R. B. Jr. 1973. Immunoglobulins of the middle ear fluid in acute otitis media: Relationships to serum immunoglobulin concentrations and bacterial cultures. *Infect. Immun.* 7:589–93

51. Bernstein, J. M., Tomasi, T. B. Jr., Ogra, P. 1974. The immunochemistry of middle ear effusions. *Arch. Otolaryngol.* 99:320–26

52. Mogi, G., Honjo, S., Maeda, S., Yoshida, T., Watanabe, N. 1974. Quantitative determination of secretory immunoglobulin A (SIgA) in middle ear effusions. *Ann. Otol. Rhinol. Laryngol.* 83:239–46

53. Ogra, P. L., Bernstein, J. M., Yurchak, A. M., Coppola, P. R., Tomasi, T. B. Jr. 1974. Characteristics of secretory immune system in human middle ear: Implications in otitis mèdia. *J. Immunol.* 112:488–95

54. Sloyer, J. L. Jr., Cate, C. C., Howie, V. M., Ploussard, J. H., Johnston, R. B. Jr. 1975. The immune response to acute otitis media in children. II. Serum and middle ear fluid antibody in otitis media due to *Haemophilus influenzae*. *J. Infect. Dis.* 132:685–88

55. Sloyer, J. L. Jr., Howie, V. M., Ploussard, J. H., Schiffman, G., Johnston, R. B. Jr. 1976. Immune response to acute otitis media: Association between middle ear fluid antibody and the clearing of clinical infection. *J. Clin. Microbiol.* 4:306–8

56. Phillips, M. J., Knight, N. J., Man-

ning, H., Abbott, A. L., Tripp, W. G. 1974. IgE and secretory otitis media. *Lancet* 2:1176–78

57. Bernstein, J. M., Reisman, R. 1974. The role of acute hypersensitivity in secretory otitis media. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* 78:120–27
58. Mogi, G., Honjo, S., Maeda, S., Yoshida, T., Watanabe, N. 1974. Immunoglobulin E (IgE) in middle ear effusions. *Ann. Otol. Rhinol. Laryngol.* 83:393–98
59. Mogi, G., Maeda, S., Yoshida, T., Watanabe, N. 1976. Radioimmunoassay of IgE in middle ear effusions. *Acta Otolaryngol.* 82:26–32
60. Lim, D. J., Lieu, Y. S., Schram, J., Birck, H. G. 1976. Immunoglobulin E in chronic middle ear effusions. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 117–23
61. Coonrod, J. D., Drennan, D. P. 1976. Pneumococcal pneumonia: Capsular polysaccharide antigenemia and antibody responses. *Ann. Intern. Med.* 84:254–60
62. Pillemer, L., Schoenberg, M. D., Blum, L., Wurz, L. 1955. Properdin system and immunity. II. Interaction of the properdin system with polysaccharides. *Science* 122:545–49
63. Winkelstein, J. A., Bocchini, J. A. Jr., Schiffman, G. 1976. The role of the capsular polysaccharide in the activation of the alternative pathway by the pneumococcus. *J. Immunol.* 116:367–70
64. Reed, W. P., Davidson, M. S., Williams, R. C. Jr., 1976. Complement system in pneumococcal infection. *Infect. Immun.* 13:1120–25
65. Bernstein, J. M., Okazaki, T., Reisman, R. E. 1976. Prostaglandins in middle ear effusions. *Arch. Otolaryngol.* 102:257–58
66. Bernstein, J. M. 1976. Biological mediators of inflammation in middle ear effusions. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 90–96
67. Lim, D. J., Liu, Y. S., Birck, H. 1976. Secretory lysozyme of the human middle ear mucosa: Immunocytochemical localization. *Ann. Otol. Rhinol. Laryngol.* 85:50–60
68. Juhn, S. K., Huff, J. S., Paparella, M. M. 1976. Certain oxidative and hydrolytic enzymes in the middle ear effusion in serous otitis media. *Ann. Otol. Rhinol. Laryngol.* 212: 119–25
69. Sade, J., Berco, E. 1976. Atelectasis and secretory otitis media. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 66–72
70. Johnson, U., Kamme, C., Laurell, A., Nilsson, N. I. 1977. C1 subcomponents in acute pneumococcal otitis media in children. *Acta Pathol. Microbiol. Scand.* 85:10–16
71. Hill, H. R., Book, L. S., Hemming, V. G., Herbst, J. J. 1977. Defective neutrophil chemotactic responses in patients with recurrent episodes of otitis media and chronic diarrhea. *Am. J. Dis. Child.* 131:433–36
72. Hussl, B., Lim, D. 1974. Experimental middle ear effusions: An immunofluorescent study. *Ann. Otol. Rhinol. Laryngol.* 83:332–42
73. Giebink, G. S., Payne, E. E., Mills, E. L., Juhn, S. K., Quie, P. G. 1976. Experimental otitis media due to *Streptococcus pneumoniae:* Immunopathogenic response in the chinchilla. *J. Infect. Dis.* 134:595–04
74. Giebink, G. S., Quie, P. G. 1977. Comparison of type 3 and 23 pneumococcal otitis media in the chinchilla model. *J. Infect. Dis.* 136: 5191–95
75. Stickler, G. B., McBean, J. B. 1954. The treatment of acute otitis media in Children. II. A second clinical trial. *J. Am. Med. Assoc.* 187:125–29
76. Rubenstein, M. M., McBean, J. B., Hedgecock, L. D., Stickler, G. B. 1965. The treatment of acute otitis media in children. III. A third clinical trial. *Am. J. Dis. Child.* 109: 308–13
77. Stickler, G. B., Rubenstein, M. M., McBean, J. B., Hedgecock, L. D., Hugstad, J. A., Griffing, T. 1967. Treatment of acute otitis media in children. IV. A fourth clinical trial. *Am. J. Dis. Child.* 114:123–30
78. Brooks, D. N. 1968. An objective method of detecting fluid in the middle ear. *Int. Audiol.* 7:280–86
79. Bluestone, C. D., Shurin, P. A. 1974. Middle ear disease in children. Pathogenesis, diagnosis, and management. *Pediatr. Clin. North Am.* 21:379–400
80. Beery, Q. C., Andrus, W. S., Bluestone, C. D., Cantekin, E. I. 1975. Tympanometric pattern classification in relation to middle ear ef-

fusions. *Ann. Otol. Rhinol. Laryngol.* 84:56–64
81. Paradise, J. L. 1977. Testing for otitis media: Diagnosis ex machina. *N. Engl. J. Med.* 296:445–48
82. Shurin, P. A., Pelton, S. I., Finkelstein, J. 1977. Tympanometry in the diagnosis of middle ear effusion. *N. Engl. J. Med.* 296:412–17
83. Paradise, J. L., Smith, C. G., Bluestone, C. D. 1976. Tympanometric detection of middle ear effusion in infants and young children. *Pediatrics* 58:198–10
84. Bluestone, C. D., Beery, Q. C., Paradise, J. L. 1973. Audiometry and tympanometry in relation to middle ear effusions in children. *Laryngoscope* 83:594–604
85. Bess, F. H., Schwartz, D. M., Redfield, N. P. 1976. Audiometric, impedance, and otoscopic findings in children with cleft palates. *Arch. Otolaryngol.* 102:465–69
86. Brooks, D. N. 1973. Hearing screening: A comparative study of an impedance method and pure tone screening. *Scand. Audio.* 2:67–72
87. Syriopoulou, V., Scheifele, D., Howie, V., Ploussard, J., Sloyer, J., Smith, A. L. 1976. Incidence of ampicillin-resistant *Hemophilus influenzae* in otitis media. *J. Pediatr.* 89:839–41
88. Bakwin, H., Jacobziner, H. 1939. Prevention of purulent otitis media in infants. *J. Pediatr.* 14:730–36
89. Heller, G. 1940. A statistical study of otitis media in children. *J. Pediatr.* 17:322–30
90. Heller, G. 1953. A statistical study of otitis media in children. II. The antibiotic era. *J. Pediatr.* 42:185–88
91. Howie, V. M., Ploussard, J. H. 1972. Efficacy of fixed combination antibiotics versus separate components in otitis media. Effectiveness of erythromycin estolate, triple sulfonamide ampicillin, erythromycin estolate-triple sulfonamide, and placebo in 280 patients with acute otitis media under two and one-half years of age. *Clin. Pediatr.* 11:205–14
92. Laxdal, O. E., Merida, J., Trefar Jones, R. H. 1970. Treatment of acute otitis media: A controlled study of 142 children. *Can. Med. Assoc. J.* 102:263–68
93. Howie, V. M., Ploussard, J. H. 1969. The "in vivo sensitivity test"—bacteriology of middle ear exudate. *Pediatrics* 44:940–44
94. Howard, J. E., Nelson, J. D., Clahsen, J., Hinton Jackson, L. 1976. Otitis media of infancy and early childhood. A double blind study of four treatment regimens. *Am. J. Dis. Child.* 130:965–70
95. Coffey, J. D. Jr. 1968. Concentration of ampicillin in exudate from acute otitis media. *J. Pediatr.* 72:693–95
96. Silverstein, H., Bernstein, J. M., Lerner, P. I. 1966. Antibiotic concentrations in middle ear effusions. *Pediatrics* 38:33–39
97. Stechenberg, B. W., Anderson, D., Chang, M. J., Dunkle, L., Wong, M., Reken, D. V., Pickering, L. K., Feigin, R. D. 1976. Cephalexin compared to ampicillin treatment of otitis media. *Pediatrics* 58:532–36
98. Howie, V. M., Ploussard, J. H., Sloyer, J. L. 1976. Natural history of otitis media. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 18–19
99. Miller, G. F. 1970. Influence of an oral decongestant on eustachian tube function in children. *J. Allergy* 45:187–93
100. Roddey, O. F. Jr., Earle, R. Jr., Haggerty, R. 1966. Myringotomy in acute otitis media. A controlled study. *J. Am. Med. Assoc.* 197:849–53
101. Crysdale, W. S. 1976. Comparative study of various ventilating tubes. *Ann. Otol. Rhinol. Laryngol.* 85: Suppl. 25, pp. 268–69
102. Kokko, E., Palva, T. 1976. Clinical results and complications of tympanostomy. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 277–79
103. Birck, H. G., Mravec, J. J. 1976. Myringostomy for middle ear effusions. Results of a two-year study. *Ann. Otol. Rhinol. Laryngol.* 85:Suppl. 25, pp. 263–67
104. Hurst, P. L. 1975. Involvement of a prepaid health plan in prospective studies on tonsillectomy and adenoidectomy. *Ann. Otol. Rhinol. Laryngol.* 84:Suppl. 19, pp. 22–24
105. Bluestone, C. D., Chairman. 1975. Workshop on tonsillectomy and adenoidectomy. State of the art and current problems. *Ann. Otol. Rhinol. Laryngol.* 84:Suppl. 19, pp. 8–14
106. McKee, W. J. E. 1963. The part

played by adenoidectomy in the combined operation of tonsillectomy with adenoidectomy: Second part of a controlled study in children. *Br. J. Prev. Soc. Med.* 17:133–40

107. Paradise, J. L. 1975. Pittsburgh tonsillectomy and adenoidectomy study: Differences from earlier studies and problems of execution. *Ann. Otol. Rhinol. Laryngol.* 84:Suppl. 19, pp. 15–19
108. Roydhouse, N. 1970. A controlled study of adenotonsillectomy. *Arch. Otolaryngol.* 92:611–16
109. Mawson, S. R., Adlington, P., Evans, M. 1967. A controlled study evaluation of adenotonsillectomy in children. *J. Laryngol. Otol.* 81:777–90
110. Ensign, P. R., Urbanich, E. M., Moran, M. 1960. Prophylaxis for otitis media in an Indian population. *Am. J. Public Health* 50:195–99
111. Perrin, J. M., Charney, E., MacWhinney, J. B. Jr., McInerny, T. K., Miller, R. L., Nazarian, L. F. 1974. Sulfisoxazole as chemoprophylaxis for recurrent otitis media. A double-blind crossover study in pediatric practice. *N. Engl. J. Med.* 291:664–67
112. Maynard, J. E., Fleshman, J. K., Tschopp, C. F. 1972. Otitis media in Alaskan Eskimo children. Prospective evaluation of chemoprophylaxis. *J. Am. Med. Assoc.* 219: 597–99
113. Howie, V. M., Ploussard, J. H., Sloyer, J. L. 1976. Immunization against recurrent otitis media. *Ann. Otol. Rhinol. Laryngol.* 85: Suppl. 25, pp. 254–58
114. Brydøy, B., Ellekejar, E. F. 1972. Otogenic meningitis: A five year study. *J. Laryngol. Otol.* 86:871–80
115. Proctor, C. A. 1966. Intracranial complications of otitic origin. *Laryngoscope* 76:288–308
116. Igarashi, M., Schuknecht, H. F. 1962. Pneumococci otitis media, meningitis, and labyrinthitis. A human temporal bone report. *Arch. Otolaryngol.* 76:126–30
117. Olmsted, R. W., Alvarez, M. C., Moroney, J. D., Eversden, M. 1964. The pattern of hearing following acute otitis media. *J. Pediatr.* 65:252–55
118. Lowe, J. F., Bamforth, J. S., Pracy, R. 1963. Acute otitis media: One year in a general practice. *Lancet* 2:1129–35
119. Fry, J., Dillane, J. B., Jones, R. F. M., Kalton, G., Andrew, E. 1969. The outcome of acute otitis media (A report to the Medical Research Council). *Br. J. Prev. Soc. Med.* 23:205–9
120. Paperella, M. M. 1976. Pathogenesis of middle ear effusions. *Ann. Otol. Rhinol. Laryngol.* 85: Suppl. 25, pp. 63–65
121. Holm, V. A., Kunze, L. H. 1969. Effect of chronic otitis media on language and speech development. *Pediatrics* 44:833–39
122. Lewis, N. 1976. Otitis media and linguistic incompetence. *Arch. Otolaryngol.* 102:387–90

Ann. Rev. Med. 1978. 29:307–18
Copyright © 1978 by Annual Reviews Inc. All rights reserved

GASTROINTESTINAL HORMONES ◆7281

James E. McGuigan, M.D.
Department of Medicine, University of Florida College of Medicine, Gainesville, Florida 32610

INTRODUCTION

Bayliss & Starling (1) identified secretin, the first known hormone, in extracts of small intestinal mucosa, more than three quarters of a century ago. During the past 10 to 15 years new information about the hormones and peptides of the gastrointestinal tract has been rapidly acquired. This has been made possible by the chemical purification of these hormones and peptides, the application of techniques for definition of their amino acid sequences, the utilization of methods for peptide synthesis, and, most recently, the use of immunological methods for cell and tissue localization and for radioimmunoassay measurement of the gastrointestinal peptides.

The substances under consideration in this discussion include a variety of peptides that have been isolated from the gastrointestinal tract, several of which can be considered gastrointestinal hormones. We examine the characteristics of these gastrointestinal peptides and their clinical significance.

GASTROINTESTINAL PEPTIDES AND HORMONES

General Considerations

A large number of polypeptides with varying biological activities have been found in extracts of tissues and in cells of the gastrointestinal tract. Among these are substances that can be considered established gastrointestinal hormones, and others that may be viewed as candidates for hormonal status. Originally, a hormonal mechanism was believed to be involved in a physiologic event during digestion if it fulfilled the following criteria. (*a*) A stimulus applied to a portion of the gastrointestinal tract provoked a biological response in a remote organ. (*b*) This biological response could be produced even when innervation between the stimulation site and the target organ had been interrupted. (*c*) The biological response could be reproduced by intravenous administration of an extract of the stimulated tissues but not by extracts of other tissues. More recently it has been required that the quantity and variety of stimulation-released hormone be sim-

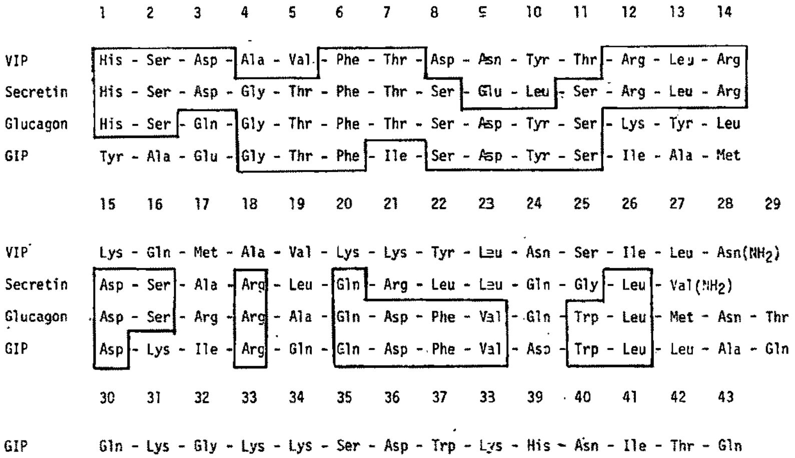
0066-4219/78/0401-0307$01.00

ilar to that required to produce the biological response by exogenous hormone administration.

Applying the above criteria we find that gastrin has achieved hormonal status, and it is probable that secretin, cholecystokinin (CCK), and gastric inhibitory peptide GIP) are also gastrointestinal hormones. Other candidates for hormone status are polypeptides purified from gut mucosa, whose amino acid composition and sequence have been defined: these include motilin, substance P, vasoactive intestinal peptide (VIP), and chymodenin. Based on their immunoreactivity, additional peptide substances have been demonstrated in gut mucosa and are referred to as glucagon-, somatostatin-, and bombesin-like materials.

The amino acid sequences of gastrin, secretin, CCK, GIP, VIP, motilin, bombesin, somatostatin, and substance P have been defined. There are regions of identical amino acid sequences between various gastrointestinal peptides. This suggests that there exist at least two families of gastrointestinal polypeptides based on these structural similarities—the secretin family and the gastrin family. Figure 1 displays the amino acid sequences and regions of amino acid identity among the peptide members of the secretin family, which include secretin, glucagon, VIP, and GIP. Multiple regions of these peptides contain identical amino acids, which probably accounts for many of the overlapping actions of these peptides.

Gastrin and CCK have been designated as members of the gastrin family. The carboxyl-terminal five amino acids in gastrin and in CCK are identical [Gly-Trp-Met-Asp-Phe (NH_2)] (2). This probably accounts for the overlapping biological actions of CCK and gastrin and for the capacity of CCK to serve as both a weak agonist to gastrin functions and also as a competitive inhibitor of gastrin action. Gastrin exists in multiple forms; the predominant form in gastric antral mucosa and in Zollinger-Ellison tumor (gastrinoma) is the heptadecapeptide amide (G-17), whereas in the blood it is big gastrin (G-34) (3–5).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
VIP	His	Ser	Asp	Ala	Val	Phe	Thr	Asp	Asn	Tyr	Thr	Arg	Leu	Arg
Secretin	His	Ser	Asp	Gly	Thr	Phe	Thr	Ser	Glu	Leu	Ser	Arg	Leu	Arg
Glucagon	His	Ser	Gln	Gly	Thr	Phe	Thr	Ser	Asp	Tyr	Ser	Lys	Tyr	Leu
GIP	Tyr	Ala	Glu	Gly	Thr	Phe	Ile	Ser	Asp	Tyr	Ser	Ile	Ala	Met

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
VIP	Lys	Gln	Met	Ala	Val	Lys	Lys	Tyr	Leu	Asn	Ser	Ile	Leu	$Asn(NH_2)$	
Secretin	Asp	Ser	Ala	Arg	Leu	Gln	Arg	Leu	Leu	Gln	Gly	Leu	$Val(NH_2)$		
Glucagon	Asp	Ser	Arg	Arg	Ala	Gln	Asp	Phe	Val	Gln	Trp	Leu	Met	Asn	Thr
GIP	Asp	Lys	Ile	Arg	Gln	Gln	Asp	Phe	Val	Asp	Trp	Leu	Leu	Ala	Gln

	30	31	32	33	34	35	36	37	38	39	40	41	42	43
GIP	Gln	Lys	Gly	Lys	Lys	Ser	Asp	Trp	Lys	His	Asn	Ile	Thr	Gln

Figure 1 Amino acid sequence of porcine peptides of the secretin family. Brackets indicate many of the regions of amino acid identity.

The gut constitutes the largest endocrine organ in the body. Unlike other endocrine organs, in which hormone-containing cells are clustered together in aggregrate tissues, endocrine cells of the gut are widely dispersed, usually singly, along the surface of the gastrointestinal mucosa. The gastrointestinal peptides are contained in cytoplasmic granules of differentiated endocrine cells that have a mucosal surface exposed to the lumen of the gastrointestinal tract. These different endocrine cells have been classified on the basis of the density, shape, and size of their secretory granules. Through use of immunological techniques (immunofluorescence and immune-peroxidase labeling), the gastrointestinal peptides have been localized in granules of specific endocrine cell types. In most instances, a single peptide has been localized to a specific cell type (6). In deviance from this pattern, however, the enterochromaffin (EC) cell contains (in addition to serotonin) in some instances substance P, and in other instances motilin, which indicates the functional heterogeneity of this cell population. H cells, in addition to VIP, also contain bombesin-like immunoreactivity. It has been suggested that most, and perhaps all, peptide hormone-producing cells are of neuroectodermal embryologic origin (6).

The distribution of the gastrointestinal peptides in the gut is variable. Maximum gastrin concentration is in the antral mucosa, with gradually decreasing concentrations of hormone in the duodenum, and small amounts in the jejunum. In contrast, substance P, VIP, somatostatin-like immunoreactivity, glucagon-like immunoreactivity, and bombesin-like immunoreactivity are present in the body of the stomach, the antrum, and throughout the small intestine and colon. GIP, motilin, CCK, and secretin are restricted to the mucosa of the duodenum, jejunum, and ileum.

Characteristics of Individual Gastrointestinal Hormones and Peptides

GASTRIN Infused gastrin induces a wide variety of actions. Its primary action is the stimulation of HCl secretion by the stomach (7, 8). It also stimulates lower esophageal sphincter pressure, relaxes the pyloric sphincter, stimulates pancreatic enzyme secretion, increases motor activity in the intestine, and modestly stimulates pancreatic bicarbonate and water secretion (9). Many of these actions occur only with supraphysiological amounts of gastrin. Gastrin increases antral motility, which is probably a physiological effect (9). Gastrin also has trophic influences on mucosal cells of the body of the stomach, small, and large intestine (10).

The development of radioimmunoassay has made possible the ready measurement of gastrin concentrations in serum, plasma, and extracts of tissues (11–13). The principal stimulus to gastrin release is feeding, influenced particularly by the protein content of the meal. Gastrin release is also stimulated by antral distention and vagal stimulation, but it is inhibited by acid within the lumen of the stomach. In man, most of the immunoreactive gastrin released in response to feeding is G-34, a little less than one third is G-17, and smaller amounts of component I gastrin and smaller gastrin variants appear in the circulation. The major gastrin species (G-34 and G-17) vary both in their rates of disappearance from the circulation and in their biological potency. The half-lives of G-17 and G-34 are approximately 6 and 40 minutes, respectively (14). Circulating concentrations of G-17 are approximately five to six times as potent as equimolar concentrations of G-34. Although more G-34 than G-17 is released, the acid secretory response to feeding is due principally to effects of G-17 because of its greater potency (14).

The primary clinical application of gastrin radioimmunoassay at present is in the detection of patients with the Zollinger-Ellison syndrome (gastrinoma) (15). These patients usually have fasting hypergastrinemia as well as exaggerated responses to intravenous calcium infusion, which increases serum gastrin concentrations by more than 50% (or by greater than 500 pg/ml) (16–18). Patients with gastrinomas exhibit paradoxical responses to intravenous secretin, with resultant increases in serum gastrin levels usually of more than 50% (or by greater than 400 pg/ml) (18–20). (In contrast, in normal individuals or patients with ordinary duodenal ulcer, intravenous secretin modestly reduces or has no effect on serum gastrin levels.)

A very small portion (less than 1%) of patients with duodenal ulcer disease may have antral G-cell hyperplasia, with gastric acid hypersecretion and marginal to slightly increased serum gastrin levels (21, 22). In these patients, intravenous secretin decreases serum gastrin, and ingestion of a protein-containing meal induces exaggerated gastrin release. This contrasts with gastrinoma patients in which serum gastrin is unchanged or increased slightly after feeding. Increased gastrin levels have also been noted in patients with achlorhydria (including those patients with pernicious anemia, carcinoma of the stomach, and common gastric mucosal atrophy) (23–25), and in patients with renal failure with the retained antrum syndrome and after massive small intestinal resection.

SECRETIN Secretin is a basic polypeptide which contains 27 amino acids and bears structural similarities to glucagon, VIP, and GIP (Figure 1). Secretin has 14 amino acids that are identical to those in glucagon. Secretin has been isolated from the mucosa of the small intestine; its concentrations are greatest in the duodenum and jejunum, and diminish with progression down the ileum.

The primary action of secretin is the stimulation of pancreatic bicarbonate and water secretion (1). Secretin, like glucagon, can inhibit smooth muscle contraction by the intestine and stomach, and is a potent inhibitor of gastric acid secretion. Secretin serves asa noncompetitive inhibitor or gastrin-mediated acid secretion and is a weak inhibitor of gastrin release (26). Release of secretin into the circulation is stimulated by hydrogen ion, with secretin release elicited with reduction of luminal pH to 4.5 or below (27). Lowering the pH below 3.0 does not further enhance pancreatic bicarbonate secretion. Secretin administration decreases lower esophageal sphincter pressure and decreases motor activity of the body and antrum of the stomach, the small intestine, and colon (26). Secretin administration stimulates pepsin secretion by the stomach and increases the tone of the pyloric sphincter. Secretin stimulation of pancreatic bicarbonate secretion is mediated by the adenylate cyclase-cyclic AMP system. Secretin augments gall bladder contraction induced by CCK, and also stimulates water and bicarbonate secretion from the liver and Brunner's glands. Secretin is a weak stimulant of pancreatic enzyme secretion. Secretin inhibits meal-stimulated gastrin secretion; however, as previously indicated, it evokes a paradoxical increase in serum gastrin in patients with Zollinger-Ellison syndrome. There is no evidence that secretin is released either by vagal stimulation or by acetylcholine administration.

In most studies it has not been possible to demonstrate secretin release into the circulation by radioimmunoassay after feeding, although increased pancreatic secretion of water and bicarbonate occur in both humans and dogs after ingestion of a meat meal (28). It is speculated that small amounts of secretin, perhaps of insufficient quantity to be detected by radioimmunoassay, are released in response to feeding and in concert with released cholecystokinin (and perhaps with VIP), stimulate the observed secretion of water and bicarbonate by the pancreas (26, 28).

CHOLECYSTOKININ (CCK) CCK is the term applied by Ivy & Oldberg in 1928 (32) to an extract of small intestinal mucosa, that, in response to lipid in the small intestine, caused gall bladder contraction. In 1943, Harper & Raper (33) suggested the name pancreozymin for a small intestinal hormone that stimulated enzyme-rich secretion by the pancreas. The hormone responsible for both CCK and pancreozymin activities is now termed cholecystokinin, or alternately, cholecystokinin-pancreozymin, and is a single polypeptide hormone containing 33 amino acids (2, 34). CCK also exists in a larger form which contains 39 amino acid residues. The carboxyl-terminal five amino acids in CCK are identical to those in gastrin. The seventh amino acid from the carboxyl-terminus is a sulfated tyrosyl residue, with the sulfation necessary for biological activity. Its primary actions are gall bladder contraction and stimulation of enzyme secretion by the pancreas. It is of interest that the carboxyl-terminal decapeptide portion of CCK is more than 10 times as potent as the intact hormone (35).

CCK induces a wide variety of biological effects such as stimulation of pepsin secretion, Brunner's gland secretion, and hepatic bicarbonate and water secretion. It enhances the activity of secretin-promoted bicarbonate secretion by the pancreas. CCK is a weak stimulant of gastric acid secretion and can inhibit competitively gastrin-mediated acid secretion by the stomach. CCK stimulation of pancreatic enzyme secretion is mediated by calcium and cyclic GMP. CCK has a trophic effect on the pancreas, and it may play a role in the induction of satiety (36).

CCK release is stimulated by hydrochloric acid, fatty acids, and amino acids. Highest concentrations of CCK are found in the jejunum, with smaller concentrations in the duodenum and the ileum (37).

Development of a satisfactorily specific and sensitive radioimmunoassay for measurement of CCK has been difficult because of the poor immunogenicity of the hormone, its lack of ready iodination, its lack of stability after iodination, and the variability in specificity of immunoassays under development. Preliminary evidence with such immunoassays suggests release of CCK in appreciable quantities into the circulation after ingestion of a meal (38). Furthermore, elevated levels of circulating CCK have been described in patients with pancreatic exocrine insufficiency (38). Further understanding of the role of CCK in pathophysiological states and in the regulation of various gastrointestinal physiological events awaits the development of more satisfactory techniques.

GASTRIC INHIBITORY PEPTIDE (GIP) GIP contains 43 amino acids, and, as a member of the secretin family, bears structural similarities to glucagon, secretin, and VIP (39). With immunofluorescent techniques and the development of antibodies to GIP, it had been possible to localize the hormone to the K endocrine cell type of the intestinal mucosa. Circulating levels of GIP can also be measured by radioimmunoassay. This peptide has been isolated from intestinal mucosa of the duodenum, jejunum, and ileum, but its highest concentration in the jejunal mucosa (40).

GIP is a powerful inhibitor of acid secretion by the stomach and, like glucagon and VIP, can stimulate secretion of intestinal juice. GIP is released into the circulation in response to feeding of either lipid, glucose, or amino acids. After glucose administration, GIP release occurs early; after lipid administration, GIP release is delayed by 2–3 hr (41). This peak response in plasma GIP concentration is associated with inhibition of pentagastrin-stimulated acid secretion. It has been proposed that GIP is the hormonal mediator of the observed mechanisms by which oral glucose ingestion induces greater insulin release than intravenous administration of the same quantity of glucose (41). Oral glucose ingestion in man produces a sustained increase in serum GIP levels capable of stimulating insulin release. The insulinotropic effect of GIP appears to be dependent upon concomitant glucose administration since fat ingestion increases serum GIP levels but not serum insulin concentrations.

By virtue of its capacity to inhibit gastric acid secretion, GIP may be viewed as one of the enterogastrones. The two major functions of this polypeptide appear to be its enterogastrone effect, manifested by inhibition of gastric acid secretion, and its insulinotropic effect in the promotion of insulin release.

VASOACTIVE INHIBITORY PEPTIDE (VIP) VIP is a polypeptide member of the secretin family that was isolated from porcine intestinal mucosa by Said & Mutt (42). It is present in the mucosa of the fundus and antrum of the stomach, the small intestine, and the colon. VIP immunoreactivity has also been found in the esophageal and rectal mucosa, the pancreas, and, more recently, in the adrenal glands and brain (6). Highest concentrations of VIP immunoreactivity in brain were found in extracts from the cerebral hemispheres (43). Immunoreactive VIP in the gut has been found in both muscular and mucosal layers. Immunofluorescence microscopy has shown that VIP was present in endocrine cell types of the gastrointestinal mucosa, as well as in nerve fibers of the myenteric plexus, lamina propria, salivary glands, and pancreas. VIP is one of a group of peptides to be found in both gut and brain; others include gastrin, substance P, somatostatin, and CCK-like immunoreactivity (6, 44, 45).

VIP induces a wide variety of biological activities when in high concentration. It produces peripheral vasodilation (46), insulin release, inhibits gastric acid secretion, stimulates hepatic glycogenolysis, stimulates pancreatic bicarbonate and water secretion (similar to secretin, but less potent), stimulates secretion of intestinal juice, increases intestinal mucosal cyclic AMP, relaxes smooth muscle, and has an inotropic action on the heart (similar to glucagon). To date, there is no proof of the VIP release with feeding. It has been suggested that VIP may be a locally acting hormone and/or serve as a neurotransmitter substance (6). Administration of VIP has been found to have no effect on basal serum gastrin levels. However, VIP, as well as GIP, has been found to suppress food-stimulated release of gastrin and gastrin-stimulated acid secretion from Heidenhain pouches in dogs (47). VIP increases bile flow, stimulates intestinal muscle contraction, and stimulates lipolysis.

VIP has been found in high concentrations in the plasma and in pancreatic islet cell tissue of patients with the pancreatic cholera or watery diarrhea, hypokalemia, and achlorhydria (WDHA) syndrome. Measurement of elevated levels of VIP in the blood of patients with this syndrome may be of value in establishing the diagnosis. Elevated plasma levels of VIP were noted in 28 of 30 patients with intractable watery diarrhea; 13 of these patients had pancreatic islet cell adenomas, 4 had islet hyperplasia, 5 had bronchogenic carcinoma, 1 had pheochromocytoma, and 1 had ganglioneuroblastoma (48). Among 21 patients with WDHA syndrome, 8 also had increased serum levels of prostaglandin E_1 and 2 had normal plasma levels of VIP (49).

ENTEROGLUCAGON Three hormonal substances have been identified with glucagon activity. The first is the well-known pancreatic glucagon (Figure 1), located in alpha cells of pancreatic islets. In addition, however, two glucagon-like peptides are extractable from gut tissue. One of these peptides is indistinguishable from pancreatic glucagon, and has a molecular weight of 3500 (50). This material may be termed gut glucagon (or enteroglucagon). It is present in the endocrine cells (indistinguishable from alpha cells of the pancreas) of the fundal mucosa of dogs (51). Unlike that in dog, enteroglucagon in man constitutes little if any (less than 3%) of total body glucagon. The second glucagon-like material in gut extracts has a molecular weight of 2900 and has less biological activity than pancreatic glucagon, with variable cross-reactivity with antibodies to pancreatic glucagon: This material may be referred to as glucagon-like immunoreactivity. Glucagon-like immunoreactivity had been found in the gastrointestinal tracts of all vertebrates thus far examined and is present in cells which resemble, but may be distinguished from, alpha cells (6).

Present evidence indicates that pancreatic glucagon and enteroglucagon are probably identical; each contains 29 amino acid residues and has a high degree of structural similarity to secretin. The amino acid composition and sequence of glucagon-like immunoreactivity are unknown. Radioimmunoassay methods can distinguish between authentic glucagon and glucagon-like immunoreactivity

(51, 52). Pancreatic glucagon decreases volume flow and enzyme secretion by the pancreas but does not reduce pancreatic bicarbonate secretion. Glucagon-like immunoreactivity increases in the blood after food ingestion; pancreatic glucagon does not. The release of insulin and glucagon-like immunoreactivity may account for the rapid decrease in blood sugar, one of the possible characteristics of patients with the dumping syndrome (53). There is no persuasive evidence of the presence of significant amounts of enteroglucagon in the tissues or in the circulation of man.

MOTILIN Motilin is a polypeptide containing 22 amino acid residues and is contained in a population of enterochromaffin cells distributed throughout the duodenum, jejunum, and ileum (54–56). Motilin has no structural similarities to any known gastrointestinal hormones or peptides.

Motilin release appears to be stimulated by alkalinization of the duodenum. When injected intravenously into dogs, motilin increases lower esophageal sphincter pressure and pepsin secretion, and delays gastric emptying. Administration of 13-norleucine-motilin (which is biologically equivalent to the native motilin) in dogs produces dose-related increases in pepsin output from gastric fistulas and a slight increase in gastric acid secretion, comparable to approximately 3% of the maximal response to pentagastrin (57). Serum gastrin levels were not affected by the administration of motilin. The role of motilin in the regulation of gastric acid secretion as well as in the control of motor functions of the intestinal tract under physiological conditions remains to be established.

SOMATOSTATIN Somatostatin, a peptide containing 14 amino acids (tetradecapeptide), was originally identified in the hypothalamus, but is also present in the mucosa of the gastrointestinal tract (58). Use of immunofluorescence located the peptide in D cells of the pancreas and upper gastrointestinal tract. Somatostatin can inhibit release of growth hormone, thyroid-stimulating hormone, glucagon, insulin, and gastrin (58). It does not inhibit release of adrenocorticotrophic hormone, luteinizing hormone, or follicle-stimulating hormone (59). In addition to inhibition of the release of gastrin, somatostatin also directly inhibits gastric acid and pepsin secretion, and motilin release (60). A somatostatin-containing islet cell tumor of the pancreas has been described (61), in which improvement of diabetes mellitus in a patient occurred following tumor resection. It is conceivable that this improvement resulted from removal of the effects of somatostatin-inhibited insulin release. The somatostatin-containing D cells of the islet, by their immediate proximity to alpha and beta cells, may inhibit release of glucagon, and/or insulin by direct within-islet action. When infused in large doses, somatostatin can inhibit gastric emptying, gastric acid secretion, gall bladder contraction, duodenal motility, and absorption of glucose and xylose by the small intestine.

SUBSTANCE P Substance P is a peptide which contains 11 amino acids (undecapeptide amide) and is found in the mucosa of the body and antrum of the stomach, the small intestine, and colon (62). Substance P immunoreactivity has also been demonstrated in pituitary tissue in high concentrations and in small concentrations in the pancreas. The major concentrations of substance P-like immunoreactivity in the gut are located in the upper portions of the jejunum and ileum. It is a candidate as a neurotransmitter (64) and was identified by von Euler & Gaddum in 1931 (63) in extracts of brain and intestinal tissue. Its principal identified actions are hypotensive manifestations in vivo and contractile effects on intestine in vitro.

CHYMODENIN Chymodenin, a large basic polypeptide with an approximate molecular weight of 5000, has been extracted from porcine duodenal mucosa (65, 66). The amino acid composition and sequence of chymodenin are yet to be determined. Administration of chymodenin to anesthetized rabbits stimulates secretion of pancreatic juice that, compared to protein secretion and other pancreatic enzymes, is enriched with chymotrypsinogen (67, 68). Results of studies with chymodenin suggest that synthesis and secretion of individual pancreatic enzymes may be modulated by specific peptide hormone regulators such as chymodenin. This argues against the previous concept of parallel pancreatic enzyme secretion by the exocrine pancreas. Further studies are required to determine the significance of chymodenin in the selective regulation of pancreatic enzyme synthesis and secretion.

PANCREATIC POLYPEPTIDE Pancreatic polypeptide contains 36 amino acids, has an approximate molecular weight of 4200, and has been extracted from mammalian and avian pancreatic tissue (69). Endocrine cells containing pancreatic polypeptide are present principally in the exocrine glandular nonislet areas of the pancreas, with smaller numbers of cells in the islets. Pancreatic polypeptide stimulates basal acid secretion, but inhibits pentagastrin-stimulated acid secretion by the stomach (70, 71). Pancreatic polypeptide in low doses enhances secretin-induced pancreatic secretion and in high doses inhibits this secretion.

Radioimmunoassay of pancreatic polypeptide has been developed. After ingestion of meals, pancreatic polypeptide in the circulation of man increases. This hormone is increased in the sera and tumors of patients with primary and metastatic tumors of the endocrine pancreas, which includes those patients with insulinomas, gastrinomas, glucagonomas, and tumors containing VIP (72). There is no known specific syndrome associated with excessive pancreatic polypeptide, nor have pancreatic tumors that contain only pancreatic polypeptide been seen.

BOMBESIN Bombesin is a polypeptide which contains 14 amino acids (tetradecapeptide) and has been isolated from the skin of certain European frogs (45). When intravenously administered in dogs, bombesin stimulates gastrin release, pancreatic and gastric acid secretion, relaxation of the choledochal sphincter, and gall bladder contraction (45). Bombesin-like immunoreactivity has been detected in the mucosa of the mammalian gastrointestinal tract (73) and may play a role in the promotion of gastrin release (45).

SUMMARY AND CONCLUSIONS

Much has been learned about the nature, localization, and biological actions of a wide assortment of peptides present in the gastrointestinal tract. Measurement of some of these peptides has been made possible by development of immunoassay techniques. In some instances, ample information has been gained about the roles of these gastrointestinal peptides in physiological events and in pathophysiological states. It is evident, however, that our knowledge of the true regulatory roles of these peptides and their participation and significance is in its earliest stages of development.

Literature Cited

1. Bayliss, W. M., Starling, E. H. 1902. The mechanism of pancreatic secretion. *J. Physiol. London* 28:325–53
2. Jorpes, J. E. 1968. Memorial lecture: the isolation and chemistry of secretin and cholecytokinin. *Gastroenterology* 55:157–64
3. Yalow, R. S., Berson, S. A. 1971. Further studies on the nature of immunoreactive gastrin in human plasma. *Gastroenterology* 60:203–14
4. Rehfeld, J. F., Stadil, F. 1973. Gel filtration studies on immunoreactive gastrin in serum from Zollinger-Ellison patients. *Gut* 14:369–73
5. Rehfeld, J. F., Stadil, F., Vikelsoe, J. 1974. Immunoreactive gastrin components in human serum. *Gut* 15:102–11
6. Pearse, A. G. E., Polak, J. M., Bloom, S. R. 1977. The newer gut hormones. *Gastroenterology* 72:746–61
7. Edkins, J. S. 1905. On the chemical mechanism of gastric secretion. *Proc. R. Soc. London* 76:376
8. Gregory, R. A., Tracy, H. J. 1964. The constitution and properties of two gastrins extracted from hog antral mucosa. *Gut* 5:103–17
9. Grossman, M. I. 1973. Spectrum of biological actions of gastrointestinal hormones. In *Frontiers in Gastrointestinal Hormone Research, Nobel Symp. 16*, ed. S. Andersson, pp. 17–28. Stockholm: Almqvist & Wiksell
10. Johnson, L. R. 1977. New aspects of the trophic action of gastrointestinal hormones. *Gastroenterology* 72:788–92
11. McGuigan, J. E., Trudeau, W. L. 1970. Studies with antibodies to gastrin: radioimmunoassay in human serum and physiological studies. *Gastroenterology* 58:139–50
12. Yalow, R. S., Berson, S. A. 1970. Radioimmunoassay of gastrin. *Gastroenterology* 58:1–14
13. Berson, S. A., Yalow, R. S. 1971. Nature of immunoreactive gastrin extracted from tissues of gastrointestinal tract. *Gastroenterology* 60:215–22
14. Walsh, J. H., Isenberg, J. I., Ansfield, J., Maxwell, V. 1976. Clearance and acid-stimulating action of human big and little gastrins in duodenal ulcer subjects. *J. Clin. Invest.* 57:1125–31
15. McGuigan, J. E., Trudeau, W. L. 1968. Immunochemical measurement of elevated levels of gastrin in the serum of patients with pancreatic tumors of the Zollinger-Ellison variety. *N. Engl. J. Med.* 278:1308–13
16. Trudeau, W. L., McGuigan, J. E. 1969. Effects of calcium on serum gastrin levels in the Zollinger-Ellison syndrome. *N. Engl. J. Med.* 281:862–66
17. Basso, N., Passaro, E. Jr. 1970. Calcium-stimulated gastric secretion in the Zollinger-Ellison syndrome. *Arch. Surg.* 101:399–402
18. Lamers, C. B. H., Van Tongeren, J. H. M. 1977. Comparative study of the value of the calcium, secretin, and meal stimulated increase in serum gastrin to the diagnosis of the Zollinger-Ellison syndrome. *Gut* 18:128–34
19. Isenberg, J. I., Walsh, J. H., Passaro, E., Moore, E. W., Grossman, M. I. 1972. Unusual effect of secretin on serum gastrin, serum calcium, and gastric acid secretion in a patient with suspected Zollinger-Ellison syndrome. *Gastroenterology* 62:626–31

20. Kolts, B. E., Herbst, C. A., McGuigan, J. E. 1974. Calcium and secretin-stimulated gastrin release in the Zollinger-Ellison syndrome. *Ann. Intern. Med.* 81:758–62
21. Straus, E., Yalow, R. S. 1974. Differential diagnosis in hyperchlorydric hypergastrinemia. *Gastroenterology* 66:867
22. Ganguli, P. C., Elder, J. B., Polak, J. M. 1974. Antral-gastrin-cell hyperplasia in peptic-ulcer disease. *Lancet* 1:1288–89
23. McGuigan, J. E., Trudeau, W. L. 1970. Serum gastrin concentrations in pernicious anemia. *N. Engl. J. Med.* 282:358–61
24. Strickland, R. G., Bhathal, P. S., Korman, M. G. 1971. Serum gastrin and the antral mucosa in atrophic gastritis. *Br. Med. J.* 4:451–53
25. McGuigan, J. E., Trudeau, W. L. 1973. Serum and tissue gastrin concentrations in patients with carcinoma of the stomach. *Gastroenterology* 64:22–25
26. Rayford, P. L., Miller, T. A., Thompson, J. C. 1976. Secretin, cholecystokinin and newer gastrointestinal hormones. *N. Engl. J. Med.* 294:1093–1101
27. Johnson, L. R., Grossman, M. I. 1968. Secretin: the enterogastrone released by acid in the duodenum. *Am. J. Physiol.* 215:885–88
28. Kolts, B. E., McGuigan, J. E. 1977. Radioimmunoassay measurement of secretin half-life in man. *Gastroenterology* 72:55–60
29. Rayford, P. L., Curtis, P. J., Fender, H. R. 1975. Radioimmunoassay measurement of disappearance half-time of secretin in dogs. *Surg. Forum* 26:385–86
30. Curtis, P. J., Rayford, P. L., Fender, H. R. 1976. Catabolism of secretin by the liver and kidneys. *Surgery* 80:259–65
31. Curtis, P. J., Rayford, P. L., Thompson, J. C. 1975. Renal extraction of circulating secretin in dogs. *Physiologist* 18:181
32. Ivy, A. C., Oldberg, E. 1928. A hormone mechanism for gall-bladder contraction and evacuation. *Am. J. Physiol.* 86:599–613
33. Harper, A. A., Raper, H. S. 1943. Pancreozymin, a stimulant of the secretion of pancreatic enzymes in extracts of the small intestine. *J. Physiol. London* 102:115–25
34. Mutt, V., Jorpes, J. E. 1968. Structure of porcine cholecystokinin-pancreozymin. I. Cleavage with thrombin and with trypsin. *Eur. J. Biochem.* 6:156–62
35. Mutt, V., Jorpes, J. E. 1973. Isolation and primary structure of cholecystokinin-pancreozymin. See Ref. 9, pp. 169–74
36. Gibbs, J., Falasco, J. D., McHugh, P. R. 1976. Cholecystokinin-decreased food intake in rhesus monkeys. *Am. J. Physiol.* 230:15–18
37. Bloom, S. R., Bryant, M. G. 1973. Distribution of radioimmunoassayable gastrin, secretin, pancreozymin and enteroglucagon in rat, dog, and baboon gut. *J. Endocrinol.* 59:44
38. Harvey, R. F., Dowsett, L., Hartog, M., Read, A. E. 1974. Radioimmunoassay of cholecystokinin-pancreozymin. *Gut* 15:690–99
39. Brown, J. C., Dryburgh, J. R. 1971. A gastric inhibitory polypeptide. II. The complete amino acid sequence. *Can. J. Biochem.* 49:867–72
40. Brown, J. C., Dryburgh, J. R., Ross, S. A. 1975. Identification and actions of gastric inhibitory polypeptide. *Recent Prog. Horm. Res.* 31:487–532
41. Brown, J. C., Dryburgh, J. R., Moccia, P. 1975. The current status of GIP. In *Gastrointestinal Hormones,* ed. J. C. Thompson, pp. 537–47. Austin: Univ. Tex. Press
42. Said, S. I., Mutt, V. 1972. Isolation from porcine-intestinal wall of vasoactive octacosapeptide related to secretin and to glucagon. *Eur. J. Biochem.* 28:199–204
43. Bryant, M. G., Polak, J. M., Modlin, I., Bloom, S. R., Albuquerque, R. H., Pearse, A. G. E. 1976. Possible dual role for vasoactive intestinal peptide as a gastrointestinal hormone and neurotransmitter substance. *Lancet* 1:991–93
44. Dockray, G. J. 1976. Immunochemical evidence of cholecystokinin-like peptides in brain. *Nature* 264:568–70
45. Rayford, P. L., Miller, T. A., Thompson, J. C. 1976. Secretin, cholecystokinin and newer gastrointestinal hormones. *N. Engl. J. Med.* 294:1157–64
46. Said, S. I., Mutt, V. 1970. Polypeptide with broad biological activity: isolation from small intestine. *Science* 169:1217–18
47. Villar, H. V., Fender, H. R., Rayford, P. L. 1975. Inhibition of gastrin release and gastric secretion by

GIP and VIP. See Ref. 41, pp. 467–73

48. Said, S. I., Faloona, G. R. 1975. Elevated plasma and tissue levels of vasoactive intestinal polypeptide in the watery-diarrhea syndrome due to pancreatic, bronchogenic and other tumors. *N. Engl. J. Med.* 293:155–60
49. Jaffe, B. M., Condon, S. 1976. Prostaglandins E and F in endocrine diarrheagenic syndromes. *Ann. Surg.* 184:516–24
50. Sasaki, H., Rubalcava, B., Srikant, C. B. 1975. Gut glucagonoid (GLI) and gut glucagon. See Ref. 41, pp. 519–28
51. Unger, R. H., Eisentraut, A. M., McCall, M. S. 1961. Glucagon antibodies and an immunoassay for glucagon. *J. Clin. Invest.* 40:1280–89
52. Valverde, I., Rigopoulou, D., Marco, J. 1970. Characterization of glucagon-like immunoreactivity (GLI). *Diabetes* 19:614–23
53. Shima, K., Kuroda, K., Matsuyama, T. 1972. Plasma glucagon and insulin responses to various sugars in gastrectomized and normal subjects. *Proc. Soc. Exp. Biol. Med.* 139:1042–48
54. Brown, J. C. 1974. Candidate hormones of the gut. IV. Motilin. *Gastroenterology* 67:734–35
55. Pearse, A. G. E., Polak, J. M., Bloom, S. R. 1974. Enterochromaffin cells of the mammalian small intestine as the source of motilin. *Virchows Arch. B* 16:111–20
56. Larsson, L. I., Sundler, F., Hakanson, R. 1975. Fluorescence histochemistry of polypeptide hormone-secreting cells in the gastrointestinal mucosa. See Ref. 41, pp. 169–95
57. Strunz, U., Domschke, W., Mitznegg, P. 1975. Analysis of the motor effects of 13-norleucine motilin on the rabbit, guinea pig, rat, and human alimentary tract in vitro. *Gastroenterology* 68:1485–91
58. Brazeua, P., Guillemin, R. 1974. Somatostatin: newcomer from the hypothalamus. *N. Engl. J. Med.* 290:963–64
59. Siler, T. M., Yen, S. S. C., Vale, W. 1974. Inhibition by somatostatin on the release of TSH induced in man by thyrotropin releasing factor. *Clin. Endocrinol.* 38:742–45
60. Bloom, S. R., Ralphs, D. N., Besser, G. M. 1975. Effect of somatostatin on motilin levels and gastric emptying. *Gut* 16:834
61. Ganda, O. P., Weir, G. C., Soeldner, J. S., Legg, M. A., Chick, W. L., Patel, Y. C., Ebeid, A. M., Gabray, K. H., Reichlin, S. 1977. "Somatostatinoma": a somatostatin-containing tumor of the endocrine pancreas. *N. Engl. J. Med.* 296:963–67
62. Chang, M. M., Leeman, S. E. 1971. Amino-acid sequence of substance P. *Nature* 232:86–87
63. Euler, U. S., Gaddum, J. H. 1931. An unidentified depressor substance in certain tissue extracts. *J. Physiol. London* 72:74–87
64. Otsuka, M., Knoishi, S., Takahashi, T. 1975. Hypothalamic substance P as a candidate for transmitter of primary afferent neurons. *Fed. Proc.* 34:1922–28
65. Adelson, J. W., Rothman, S. S. 1974. Candidate hormones of the gut. II. Chymodenin. *Gastroenterology* 67:731–33
66. Adelson, J. W. 1975. Chymodenin: an overview. See Ref. 41, pp. 553–74
67. Adelson, J. W., Rothman, S. S. 1974. Selective pancreatic enzyme secretion due to a new peptide called chymodenin. *Science* 183: 1087–89
68. Adelson, J. W., Rothman, S. S. 1975. Chymodenin, a duodenal peptide: specific stimulation of chymotrypsinogen secretion. *Am. J. Physiol.* 229:1680–86
69. Kimmel, J. R., Pollock, H. G., Hazelwood, R. L. 1968. Isolation and characterization of chicken insulin. *Endocrinology* 83:1323–30
70. Lin, T. M., Chance, R. E. 1972. Spectrum gastrointestinal actions of a new bovine pancreas polypeptide (BPP). *Gastroenterology* 62:852
71. Lin, T. M., Chance, R., Evans, D. 1973. Stimulatory and inhibitory actions of a bovine pancreatic peptide (BPP) on gastric and pancreatic secretions of dogs. *Gastroenterology* 64:865
72. Polak, J. M., Bloom, S. R., Adrian, T. E. 1976. Pancreatic polypeptide in insulinomas, gastrinomas, vipomas, and glucagonomas. *Lancet* 1:328–30
73. Lechago, J., Holmquist, A. L., Walsh, J. H. Immunofluorescence localization of bombesin cells of amphibian upper gastrointestinal mucosa. *Gastroenterology*. In press

Ann. Rev. Med. 1978. 29:319–25
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CHRONIC BACTEREMIA AND NEPHRITIS

◆7282

Youngki Kim, M.D. and Alfred F. Michael, M.D.
Department of Pediatrics, University of Minnesota Medical School, Minneapolis, Minnesota 55455

The occurrence of glomerulonephritis has been recognized in a variety of systemic infections caused by bacteria, viruses, fungi, rickettsiae and protozoa. Evidence accumulated in recent years suggests that most, if not all, of these glomerulonephritides are caused by the glomerular deposition of circulating immune complexes as a result of the host's immune response to bacterial antigens. Although host factors are not entirely understood, the immune responsiveness of the host (which determines the amount, type, and class of the specific antibody, and which in lower animals is genetically determined) plays a critical role in the pathogenesis of these nephritides.

A number of bacterial agents have been implicated in the development of glomerulonephritis in human beings (Table 1). This review includes a discussion of glomerulonephritis associated with bacterial endocarditis, ventriculoatrial shunt infections, and syphilis.

BACTERIAL ENDOCARDITIS

Renal lesions associated with bacterial endocarditis are of two main types, embolic and nephritic. Embolism occurs quite frequently, sometimes in coexistence with nephritis. Kidneys frequently have a "flea-bitten" appearance due to subcapsular petechial hemorrhages. Some of the glomeruli show intracapillary thrombosis and necrosis with or without local cellular proliferation, while others are entirely normal. Areas of infarct may also be present. Löhlein (1) and Baehr (2) found bacteria in renal lesions of these patients and suggested the importance of bacterial embolization. Most of the subsequent studies, however, failed to demonstrate any organisms, which indicates that the renal lesions are most probably the consequence of embolization of friable vegetations of heart valves. Other than hematuria, which may sometimes be macroscopic, these patients usually remain asymptomatic and renal function is rarely impaired (3, 4). Flank pain may be present with extensive renal infarction.

0066-4219/78/0401-0319$01.00

Table 1 Bacterial infection and nephritis

Clinical condition	Renal pathology	Infectious agents	References
Endocarditis	Proliferative glomerulonephritis	*Streptococcus viridans*	1, 2, 4, 35
		Staphylococcus aureus	5, 36
		Enterococcus	7
		Pneumococcus	37
		Hemophilus influenzae	38
		Gonococcus	4, 37
		Bacteroides	39
		Brucella suis	40
		Propionibacterium acnes	41
Infected ventriculoatrial shunt	Membranoproliferative glomerulonephritis	*Staphylococcus albus*	11–14, 16–18
		Corynebacterium bovis	15
		Staphylococcus aureus	16
		Listeria monocytogenes	13
Syphilis (congenital and acquired)	Proliferative glomerulonephritis	*Treponema pallidum*	19–34, 42, 43
Typhoid fever	Mesangioproliferative glomerulonephritis Interstitial nephritis	*Salmonella typhi*	44, 45
Leprosy	Proliferative glomerulonephritis	*Mycobacterium leprae*	46
Tuberculosis	Proliferative glomerulonephritis	*Mycobacterium tuberculosis*	47
Osteomyelitis	Membranoproliferative glomerulonephritis	*Staphylococcus aureus*	48
Abdominal abscess	Membranoproliferative glomerulonephritis Focal proliferative glomerulonephritis Extracapillary glomerulonephritis	*Staphylococcus aureus* *Moraxella alcaligenes* *Pseudomonas aeruginosa* *Escherichia coli* *Proteus mirabilis*	49
Pneumonia	Membranoproliferative glomerulonephritis	*Pneumococcus*	50

On the other hand, the nephritic lesion is frequently associated with impairment of renal function, which generally correlates with the severity of morphologic findings (4, 5). Hematuria and proteinuria are usually present, although nephrotic syndrome is a rare occurrence. Rheumatoid factor and decreased level of serum complement are frequently present. Light microscopy shows diffuse or focal proliferative glomerular changes, indistinguishable from other types of acute proliferative glomerulonephritis (5). Granular deposits of IgG, IgM, IgA, and C3 are found along glomerular basement membrane (GBM) and/or in mesangium by immunofluorescence microscopy (5–7). Electron microscopy reveals electron-dense deposits along GBM (subepithelial, intramembranous, or subendothelial) and occasionally in mesangium (5). These morphological findings are consistent with immune-complex–mediated renal injury not unlike those seen in lupus nephropathy or experimental serum sickness. Williams et al (8) reported the appearance of rheumatoid factor and hypocomplementemia in these patients and proposed that the renal injury was immune-mediated. Levy et al (7) were able to elute antibody from the kidney of a patient with bacterial endocarditis that specifically reacted with bacteria from the patient's blood. Arnold et al (9) succeeded in producing diffuse glomerulonephritis in rabbits with experimentally induced bacterial endocarditis. Morphological and serological findings of these rabbits were very similar to those seen in human beings with endocarditis-associated glomerulonephritis. Furthermore, demonstration of circulating immune complexes in patients with bacterial endocarditis (10) substantiates the concept that endocarditis-associated glomerulonephritis is due to immune complexes formed as a result of host immune response to bacterial antigens. Eradication of organisms in these patients usually results in complete healing of the renal lesions with reversal of renal dysfunction and hypocomplementemia (5).

SHUNT NEPHRITIS

In 1965 Black, et al (11) first reported two patients with nephrotic syndrome associated with bacteremia after ventriculoatrial shunt operations for hydrocephalus. They suggested an immunological pathogenesis as a consequence of chronic bacteremia, a concept supported by subsequent studies that demonstrated immunoproteins and electron-dense deposits in the kidneys by immunofluorescence and electron microscopy (12, 13). Strife et al (13) have shown the presence of infecting bacterial antigens and specific antibodies in the cryoglobulins obtained from patients with shunt nephritis, evidence that the cryoglobulins contained immune complexes. The decreased levels of serum complement and the demonstration of bacterial antigens on the diseased glomeruli (12, 14) also support this concept. Renal manifestations include hematuria (gross or microscopic), proteinuria, and azotemia. Hypertension is not uncommon, and nephrotic syndrome is present in approximately one half of the patients. The levels of serum complement are usually decreased (12–16). Rheumatoid factor and cryoglobulins may be present in the serum (12–14). Blood cultures are invariably positive, the most common organism being *Staphylococcus albus*. Light microscopy shows changes similar to those seen in idiopathic membranoproliferative glomerulonephritis (12, 13); diffuse involvement of glomeruli with lobulation, increase in mesangial matrix and cellularity, GBM thickening with splitting, and occasional infiltration of neutrophils. Immunofluorescence microscopy (12, 13) demonstrates granular deposition of IgG, IgM, C1q, C3, and C4 in the mesangium and along the GBM, especially of the peripheral capillary loops. Properdin and fibrin may be found in the same location. Electron microscopy (12, 13) reveals an increase in mesangial matrix and cellularity with interposition of mesangial cell cytoplasmic processes. Electron-dense deposits are found in the subendothelial and mesangial areas. The prognosis is good once bacteriological cure is achieved by removal of the shunt and antimicrobial therapy (11, 12, 16–18). Renal manifestations usually abate promptly, except in a few patients who have residual low-grade proteinuria (12, 16).

SYPHILITIC NEPHRITIS

Although very uncommon, renal involvement associated with syphilis has long been recognized (Table 1). Two clinical expressions of syphilitic glomerulonephritis may be observed (19): in some patients marked proteinuria and nephrotic syndrome predominate, whereas in others "nephritic symptoms" (hematuria, hypertension, azotemia) are apparent.

In acquired syphilis, the nephrotic syndrome is by far the most common renal manifestation with marked proteinuria, edema, hypoalbuminemia, and hypercholesterolemia; significant hematuria, azotemia, or hypertension is usually absent (19–25). A nephritic picture occurs infrequently and is characterized by hematuria (gross or microscopic), proteinuria, azotemia, and hypertension with or without nephrotic syndrome (19). The level of serum complement is usually normal (20, 22, 24, 25). The pathological pictures of both types are very similar (20, 23–27). On light microscopy, glomeruli are hypercellular mainly due to endothelial and mesangial cell proliferation with an increase in mesangial matrix. The GBM is thickened, frequently in a segmental distribution. Focal areas of interstitial edema with infiltration of neutrophils and mononuclear cells are present. Occasionally renal tissue may appear to be normal (22). Immunofluorescence microscopy demonstrates granular deposition of IgG and C3 along the GBM and occasionally in mesangium. IgM may also be detected in a mesangial localization, whereas IgA is usually not observed (20, 22, 24–27). Electron microscopy shows irregular thickening of the GBM with electron-dense deposits primarily in subepithelial area, less frequently in intramembranous, subendothelial, or mesangial area. Fusion of epithelial cell foot-processes, and increases in mesangial matrix and cellularity may be seen (20, 23–27). Most patients respond promptly (usually within 2–4 weeks) to antisyphilitic treatment (20, 25), although in some instances spontaneous remission may occur prior to the initiation of therapy (22, 23, 25).

In congenital syphilis, a mixed variety of nephritic and nephrotic manifestations occur. Patients usually develop nephrotic syndrome with significant hematuria, azotemia, and hypertension (28–32). In contrast to the glomerulonephritis in acquired syphilis, the levels of serum complement are frequently decreased (31–32a). Histopathological, immunopathological, and ultrastructural findings are similar to those seen in acquired syphilis (26, 28, 29, 31–34), except for a more prominent proliferation of glomerular epithelial cells (28, 29). Reversal of renal manifestations of congenital syphilis is similar to that of acquired syphilis, although no spontaneous remission has ever been reported (26, 28, 29, 31, 32). The immunopathological and ultrastructural findings in the kidney in congenital and acquired syphilis clearly indicate an immune pathogenesis. Recent demonstration of antitreponemal antibody and treponemal antigen in these glomeruli further supports the concept of immune-complex–mediated renal injury (24, 26, 27), although circulating immune complexes have not been demonstrated in these patients.

CONCLUSION

Evidence that immune mechanisms play a decisive role in the pathogenesis of various forms of glomerulonephritis associated with bacterial infection includes: the demonstration of glomerular deposits of immune globulin, complement components, and in some instances bacterial antigen; the decreased serum levels of classical and alternative complement pathway components; the presence of immune complexes in the serum using sensitive assays in some diseases; and correlative studies that demonstrate morphologic and immunopathologic similarity to experimental immune-complex disease. On theoretical grounds, any organism that is able to infect an immunocompetent host over a relatively long period of time, which leads to prolonged bacteremia or antigenemia, has the potential to induce immune-complex disease. The immune-response variables of the patient may be an important determinant of this process, although this has not been evaluated. Bacteria-related immune-complex disease is a relatively uncommon cause of renal disease when compared with the many forms of idiopathic glomerulonephritis that affect human beings. Although attempts to identify an infectious agent in the latter group of diseases have been unsuccessful, the morphologic and immunopathologic similarity to bacteria-induced glomerular disease suggests chronic exposure to some as yet unidentifiable agent. Available information indicates that in most forms of glomerulonephritis the kidney is hurt as an innocent bystander in an immunologic process. In this context, chronic bacterial infections reside at a site remote from the kidney, and it is the immune response to the organism and its antigen(s) that ultimately leads to the development of glomerulonephritis.

Literature Cited

1. Löhlein, M. 1910. Ueber hamorrhagische nierenaffektionen bei chronischer ulzeroser endokarditis. (Embolische nichteiterige herdnephritis). *Med. Klin. (Munich)* 10:375–79
2. Baehr, G. 1912. Glomerular lesions of subacute bacterial endocarditis. *J. Exp. Med.* 15:330–47
3. Lerner, P. I., Weinstein, L. 1966. Infective endocarditis in the antibiotic era. *N. Engl. J. Med.* 274: 259–66
4. Villarreal, H., Sokoloff, L. 1950. The occurrence of renal insufficiency in subacute bacterial endocarditis. *Am. J. Med. Sci.* 220:655–61
5. Gutman, R. A., Stiker, G. E., Gililand, B. C., Cutler, R. E. 1972. The immune complex glomerunonephritis of bacterial endocarditis. *Medicine* 51:1–25
6. Keslin, M. H., Messner, R. P., Williams, R. C. 1973. Glomerulonephritis with subacute bacterial endocarditis. Immunofluorescent studies. *Arch. Intern. Med.* 132:578–81
7. Levy, R. L., Hong, R. 1973. The immune nature of subacute bacterial endocarditis (SBE) nephritis. *Am. J. Med.* 54:645–52
8. Williams, R. C., Kunkel, H. C. 1962. Rheumatoid factor, comple-

ment, and conglutinin aberrations in patients with subacute bacterial endocarditis. *J. Clin. Invest.* 41:666–75

9. Arnold, S. B., Valone, J. A., Askenase, P. W., Kashgarian, M., Freedman, L. R. 1975. Diffuse glomerulonephritis in rabbits with *Streptococcus viridans* endocarditis. *Lab. Invest.* 32:681–89
10. Bayer, A. S., Theofilopoulos, A. N., Eisenberg, R., Dixon, F. J., Guze, L. B. 1976. Circulating immune complexes in infective endocarditis. *N. Engl. J. Med.* 295: 1500–5
11. Black, J. A., Challacombe, D. N., Ockenden, B. G. 1965. Nephrotic syndrome associated with bacteraemia after shunt operations for hydrocephalus. *Lancet* 2:921–24
12. Dobrin, R. S., Day, N. K., Quie, P. G., Moore, H. L., Vernier, R. L., Michael, A. F., Fish, A. J. 1975. The role of complement, immunoglobulin and bacterial antigen in coagulase-negative staphylococcal shunt nephritis. *Am. J. Med.* 59:660–73
13. Strife, C. F., McDonald, B. M., Ruley, E. J., McAdams, A. J., West, C. D. 1976. Shunt nephritis: the nature of the serum cryoglobulins and their relation to the complement profile. *J. Pediatr.* 88:403–13
14. Kaufman, D. B., McIntosh, R. 1971. The pathogenesis of the renal lesion in a patient with streptococcal disease, infected ventriculoatrial shunt, cryoglobulinemia and nephritis. *Am. J. Med.* 50:262–68
15. Bolton, W. K., Sande, M. A., Normansell, D. E., Sturgill, B. C., Westervelt, F. B. 1975. Ventriculojugular shunt nephritis with *Corynebacterium bovis*. Successful therapy with antibiotics. *Am. J. Med.* 59:417–23
16. Stickler, G. B., Shin, M. H., Burke, E. C., Holley, K. E., Miller, R. H., Segar, W. E. 1968. Diffuse glomerulonephritis associated with infected ventriculoatrial shunt. *N. Engl. J. Med.* 279:1077–82
17. Rames, L., Wise, B., Goodman, J. R., Piel, C. F. 1970. Renal disease with *Staphylococcus albus* bacteremia. A complication in ventriculoatrial shunts. *J. Am. Med. Assoc.* 212:1671–77
18. Moncrieff, M. W., Glasgow, E. F., Arthur, L. J. H., Hargreaves, H. M. 1973. Glomerulonephritis associated with *Staphylococcus albus* in a Spitz-Holter valve. *Arch. Dis. Child.* 48:69–72
19. Thomas, E. W., Schur, M. 1946. Clinical nephropathies in early syphilis. *Arch. Intern. Med.* 78:679–86
20. Bhorade, M. S., Carag, H. B., Lee, H. J., Potter, E. V., Dunea, G. 1971. Nephropathy of secondary syphilis. A clinical and pathological spectrum. *J. Am. Med. Assoc.* 216:1159–66
21. Robins, D. E., Ladd, A. T. 1962. Acute syphilitic nephrosis. Case report and review of the literature. *Am. J. Med.* 32:817–21
22. Sterzel, R. B., Krause, P. H., Zobl, H., Kuhn, K. 1974. Acute syphilitic nephrosis: a transient glomerular immunopathy. *Clin. Nephrol.* 2:164–68
23. Falls, W. F., Ford, K. L., Ashworth, C. T., Carter, N. W. 1965. The nephrotic syndrome in secondary syphilis. Report of a case with renal biopsy findings. *Ann. Intern. Med.* 63:1047–58
24. Gamble, C. N., Reardan, J. B. 1975. Immunopathogenesis of syphilitic glomerulonephritis. Elution of antitreponemal antibody from glomerular immune-complex deposits. *N. Engl. J. Med.* 292:449–54
25. Braunstein, G. D., Lewis, E. J., Galvanek, E. G., Hamilton, A., Bell, W. R. 1970. The nephrotic syndrome associated with secondary syphilis. An immune deposit disease. *Am. J. Med.* 48:643–48
26. O'Regan, S., de Chadarevian, J.-P., Rishikof, J. R., Drummond, K. N. 1976. Treponemal antigens in congenital and acquired syphilitic nephritis. *Ann. Intern. Med.* 85:325–27
27. Tourville, D. R., Byrd, L. H., Kim, D. U., Zajd, D., Lee, I., Reichman, L. B., Baskin, S. 1976. Treponemal antigen in immunopathogenesis of syphilitic glomerulonephritis. *Am. J. Pathol.* 82:479–88
28. Kaschula, R. O. C., Uys, C. J., Kuijten, R. H., Dale, J. R. P., Wiggelinkhuizen, J. 1974. Nephrotic syndrome of congenital syphilis. Biopsy studies in four cases. *Arch. Pathol.* 97:289–96

29. Hill, L. L., Singer, D. B., Falletta, J., Stasney, R. 1972. The nephrotic syndrome in congenital syphilis: An immunopathy. *Pediatrics* 49:260–66
30. Pollner, P. 1966. Nephrotic syndrome associated with congenital syphilis. *J. Am. Med. Assoc.* 198: 173–76
31. Wiggelinkhuizen, J., Kaschula, R. O. C., Uys, C. J., Kuijten, R. H., Dale, J. 1973. Congenital syphilis and glomerulonephritis with evidence for immune pathogenesis. *Arch. Dis. Child.* 48:375–81
32. Yuceoglu, A. M., Tresser, G., Wasserman, E., Lange, K. 1974. The glomerulopathy of congenital syphilis. A curable immune-deposit disease. *J. Am. Med. Assoc.* 229: 1085–89
32a. Zelazko, M., Feldman, G. 1976. Behavior of the complement system in the nephropathy of congenital syphilis. *J. Pediatr.* 88:359
33. Taitz, L. S., Isaacson, C., Stein, H. 1961. Acute nephritis associated with congenital syphilis. *Br. Med. J.* 2:152–53
34. Kaplan, B. S., Wiglesworth, F. W., Marks, M. I., Drummond, K. N. 1972. The glomerulopathy of congenital syphilis—an immune deposit disease. *J. Pediatr.* 81:1154–56
35. Morel-Maroger, L., Sraer, J.- D., Herreman, G., Godeau, P. 1972. Kidney in subacute endocarditis. Pathological and immunofluorescence findings. *Arch. Pathol.* 94:205–13
36. Tu, W. H., Shearn, M. A., Lee, J. C. 1969. Acute diffuse glomerulonephritis in acute staphylococcal endocarditis. *Ann. Intern. Med.* 71:335–41
37. Bain, R. C., Edward, J. E., Scheifley, C. H., Geraci, J. E. 1958. Right-sided bacterial endocarditis and endarteritis. A clinical and pathologic study. *Am. J. Med.* 24:98–110
38. Baehr, G. 1931. Renal complications of endocarditis. *Trans. Assoc. Am. Physicians* 46:87–95
39. Case records of the Massachusetts General Hospital. 1974. *N. Engl. J. Med.* 291:242–49
40. DeGowin, E. L., Carter, J. R., Borts, I. H. 1945. A case of infection with *Brucella suis*, causing endocarditis and nephritis; death from rupture of mycotic aneurysm. *Am. Heart J.* 30:77–87
41. Beeler, B. A., Crowder, J. G., Smith, J. W., White, A. 1976. Propionibacterium acnes: pathogen in central nervous system shunt infection. Report of three cases including immune complex glomerulonephritis. *Am. J. Med.* 61:935–38
42. Hellier, M. D., Webster, A. D. B., Risinger, A. J. M. F. 1971. Nephrotic syndrome: A complication of secondary syphilis. *Br. Med. J.* 4:404–5
43. Suskind, R., Winkelstein, J. A., Spear, G. A. 1973. Nephrotic syndrome in congenital syphilis. *Arch. Dis. Child.* 48:237–39
44. Faierman, D., Ross, F. A., Seckler, S. G. 1972. Typhoid fever complicated by hepatitis, nephritis, and thrombocytopenia. *J. Am. Med. Assoc.* 221:60–61
45. Sitprija, V., Pipatanagul, V., Boonpucknavig, V., Boonpucknavig, S. 1974. Glomerulitis in typhoid fever. *Ann. Intern. Med.* 81:210–13
46. Iveson, J. M. I., McDougall, A. C., Leathem, A. J., Harris, H. J. 1975. Lepromatous leprosy presenting with polyarthritis, myositis, and immune-complex glomerulonephritis. *Br. Med. J.* 3:619–21
47. Berman, L. B., Antonovych, T. T., Duke, J. 1960. Glomerular abnormalities in tuberculosis. *Arch. Pathol.* 69:278–285
48. Boonshaft, B., Maher, J. F., Schreiner, G. E. 1970. Nephrotic syndrome associated with osteomyelitis without secondary amyloidosis. *Arch. Intern. Med.* 125:322–37
49. Beaufils, M., Morel-Maroger, L., Sraer, J.- D., Kanfer, A., Kourilsky, O., Richet, G. 1976. Acute renal failure of glomerular origin during visceral abscesses. *N. Engl. J. Med.* 295:185–89
50. Hyman, L. R., Jens, E. H., Hill, G. S., Zimmerman, S. W., Burkholder, P. M. 1975. Alternate C3 pathway activation in pneumococcal glomerulonephritis. *Am. J. Med.* 58:810–14

Ann. Rev. Med. 1978. 29:327–42
Copyright © 1978 by Annual Reviews Inc. All rights reserved

OSTEOMALACIA AND DISORDERS OF VITAMIN D METABOLISM

◆7283

Joel F. Habener, M.D.[1] *and Jane E. Mahaffey, M.D.*[2]
Endocrine Unit, Massachusetts General Hospital and Department of Medicine,
Harvard Medical School, Boston, Massachusetts 02114

INTRODUCTION

During the past decade, understanding of the pathophysiological processes that underlie the development of certain of the osteomalacia syndromes has been greatly accelerated by the elucidation of the pathways of the metabolism and the mechanism of action of vitamin D.

Investigations have revealed that the classic vitamin D compounds cholecalciferol (D_3) and ergocalciferol (D_2) are themselves metabolically inactive at target tissues, and must first undergo a series of hydroxylation reactions in the liver and kidney, which lead to the formation of the final active hydroxylated compounds. It appears that defects in this pathway of metabolism may explain many of the vitamin D–resistant diseases.

This brief review focuses principally on the pathogenesis of osteomalacia as it relates to new information concerning specific inherited or acquired defects in the metabolism of vitamin D; it also indicates how this newer information may be applied to formulate rational approaches to the management of patients with osteomalacia. Several excellent reviews on the subjects of osteomalacia (1–3) and vitamin D metabolism (4–8) were added to the literature in the past decade.

DEFINITION

Osteomalacia is a disorder with diverse causes (Table 1) characterized by a failure of normal mineralization of bone. In children, the disease takes the form of rickets, in which undermineralization of both cartilage and bone occurs, with resultant retardation of growth and the development of characteristic skeletal deformities. In adults in whom bone growth has ceased, the manifestations of

---

[1] Investigator, Howard Hughes Medical Institute

[2] Teaching and Research Scholar of the American College of Physicians

0066-4219/78/0401-0327$01.00

osteomalacia are more subtle. In its mild form, the disease may remain asymptomatic for years. In more advanced forms, the disease presents with the typical findings of diffuse bone pain, muscle weakness, hypophosphatemia, hyperphosphatasia, low to low-normal serum calcium, diminished urinary calcium, a roentgenographic appearance of diffuse osteopenia with or without characteristic pseudofractures, and a bone biopsy that reveals the pathognomonic widening of the osteoid seams.

ETIOLOGY

The defect in mineralization that occurs in osteomalacia may arise from a combination of deficiencies of vitamin D or phosphate, or a defect in production of alkaline phosphatase or bone matrix. Calcium deficiency per se does not lead to osteomalacia, although vitamin D deficiency is usually manifested by deficiencies of both calcium and phosphate. In addition, the development of osteomalacia is frequently accompanied by some degree of secondary hyperparathyroidism, presumably induced by the decreased calcium concentration in the extracellular fluid.

Osteomalacia has traditionally been classified as either responsive or resistant to treatment with vitamin D. The vitamin D–responsive disorders include those due to a deficiency of the parent vitamin D compounds ergocalciferol (vitamin D_2) or cholecalciferol (vitamin D_3) that results from inadequate dietary intake, intestinal malabsorption, or inadequate exposure to sunlight. The group of vitamin D–resistant disorders results from diverse etiologies such as familial or acquired renal-tubular phosphate wasting (phosphate diabetes), hereditary deficiency of alkaline phosphatase (hypophosphatasia), defective matrix formation (osteogenesis imperfecta), drugs that inhibit mineralization (fluorides, diphosphonates), and what appear to be specific defects in the metabolic transformation of vitamin D to its active form (hereditary vitamin D–dependent osteomalacia, renal osteodystrophy, anticonvulsant osteomalacia). The establishment of this latter category, defects in the metabolism of vitamin D, resulted from the elucidation of the metabolic pathways involved in the conversion of vitamin D to the active forms, the biochemical mechanisms of its actions on the target tissues (intestine and bone), and the factors responsible for the regulation of production of the active forms of the vitamin.

CHEMISTRY AND METABOLISM OF VITAMIN D

Although osteomalacia and rickets were recognized as distinct clinical syndromes in the 1800s (9), the cause of the disorders was not defined until 1917 when Hess & Unger (10) described the prevention of rickets in children by the administration of cod liver oil or food that had been irradiated with ultraviolet light. In 1919 Huldschinsky (11) demonstrated the cure of rickets in children by external irradiation with ultraviolet light. Isolation of the fat-soluble factors responsible for the cure of osteomalacia showed them to be sterols, ergocalciferol (vitamin D_2) and cholecalciferol (vitamin D_3) (12, 13). Both of these compounds are derived from cholesterol precursors by ultraviolet irradiation that opens the ring structure between C9 and C10 (Figure 1). Vitamin D_2, the principal dietary (exogenous) form of the vitamin, is prepared by irradiation of ergosterol. Vitamin D_3, the endogenous form, arises from irradiation of 7-dehydrocholesterol (Pro-D_3) in the skin. Once formed, the vitamins are transported in the blood stream by specific proteins. The availability of tritiated vitamin D_3 in the 1960s made possible a series of investigations of the metabolism of the vitamin (14). These investigations rapidly led to the discovery that both vitamins must undergo two separate enzymatic hydroxylations before they

Table 1 Classification of the osteomalacia syndromes

- I. Deficiency of vitamin D activity
  - A. Inadequate availability of vitamin D
    - 1. Dietary deficiency or lack of sun exposure
    - 2. Malabsorption syndromes (postgastrectomy; pancreatic and small-bowel disease; cholestyramine)
  - B. Defects in the metabolism of vitamin D
    - 1. 25-hydroxylation
      - a. Anticonvulsant drugs (phenobarbital, diphenylhydantoin)
      - b. Liver disease (primary biliary cirrhosis)
    - 2. 1-hydroxylation
      - a. Hereditary vitamin D-dependent rickets ("pseudo-vitamin D-deficiency rickets")
      - b. Hypoparathyroidism[a]
      - c. Pseudohypoparathyroidism[a]
      - d. Nephrectomy or severe renal failure
      - e. Oncogenic osteomalacia
  - C. Impairment of actions of $1,25(OH)_2D$ on target tissues
    - 1. Anticonvulsant drugs
    - 2. Uremia
    - 3. ? Corticosteroids
- II. Phosphate deficiency
  - A. Experimentally induced
  - B. Renal-tubular defects (phosphate diabetes)
    - 1. Familial hypophosphatemic rickets
    - 2. Acquired hypophosphatemia
      - a. Sporadic adult onset without known cause
      - b. Fanconi syndrome (e.g. heavy-metal poisoning, Wilson's disease)
- III. Hypophosphatasia
- IV. Defects in formation of bone matrix (osteogenesis imperfecta)
- V. Other
  - A. Fluorides
  - B. Diphosphonates
  - C. Hyperparathyroidism post-parathyroidectomy

[a] Hypoparathyroidism and pseudohypoparathyroidism only rarely are accompanied by osteomalacia.

can function at the target cells (4–7). The first hydroxylation occurs in the liver (15) and results in the formation of 25-hydroxyvitamin D. 25-Hydroxyvitamin D is carried in the blood on the specific carrier protein to the kidney where it is converted to 1,25-dihydroxyvitamin D by a 25-hydroxyvitamin D-1α-hydroxylase (16, 17). This $1,25(OH)_2D$ is transported (again bound to the transport protein in blood) to the target cells in intestine, bone, and probably also kidney.

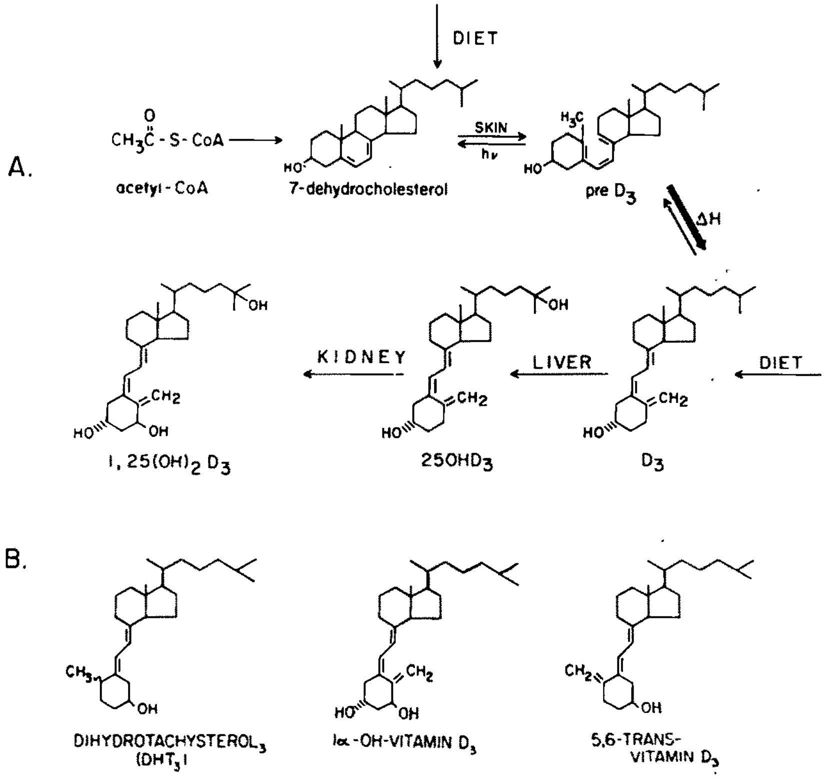


Figure 1 A. Metabolic pathway of cholecalciferol (vitamin D_3). 7-Dehydrocholesterol (provitamin D), whether synthesized from acetyl-CoA in the body or taken in the diet, is converted in the skin under the stimulus of ultraviolet irradiation to cholecalciferol (vitamin D_3) via a short-lived intermediate 9,10-seco-steroid with a 5,6-cis-triene conformation (previtamin D_3) (72). Vitamin D_3 is converted in the liver to 25-hydroxycalciferol [$25(OH)D_3$], which in turn is converted in the kidney to 1,25-dihydroxycholecalciferol [$1,25(OH)_2D_3$] by specific hydroxylases. The latter 1-hydroxylated metabolite is believed to be the biologically active form of the vitamin. A similar pathway is believed to exist for ergocalciferol (vitamin D_2), which differs structurally from cholecalciferol (vitamin D_3) in that it has a 9-carbon rather than an 8-carbon side chain. B. Analogues of vitamin D_3 that are highly effective in the treatment of patients with metabolic blocks at the level of the 1-hydroxylation. All analogues have hydroxyl groups in the 1-position of the A ring.

The evidence indicates that the mechanism of action of 1,25-dihydroxyvitamin D is probably similar to that of other steroid hormones. The $1,25(OH)_2D$ acts on the intestinal target cells by binding to cytosolic receptor proteins, and the steroid-receptor complex is carried to the nucleus of the cell where synthesis of messenger RNA for a calcium-binding or calcium-transport protein is stimulated (8). The net result is an active transport of calcium into the extracellular fluid. Similar transport systems may also exist for phosphate (7).

As a result of the identification of the hydroxylated metabolites of vitamin D, specific competitive-binding assays have been developed using synthetic tritiated metabolites and either plasma-, cytosolic-, or nuclear-binding proteins (18–20). Applications of the assays for $1,25(OH)_2D$ have shown that nephrectomy results in a loss of detectable blood levels of the dihydroxylated sterol (19, 20). Moreover, the conversion of 25-hydroxyvitamin D to $1,25(OH)_2D$ in the kidney is a regulated process (4–7) (Figure 2). The main factors that enhance the conversion of 25(OH)D to $1,25(OH)_2D$ are hypocalcemia (21), parathyroid hormone (PTH) (22–24), and hypophosphatemia (25, 26). It is likely that reduction in renal-cell phosphate levels may be the final step in the regulatory process, and that effects of hypocalcemia are mediated through the actions of PTH; these effects are in turn expressed via the PTH action of lowering renal-cell phosphate.

The question of significant physiological regulation of the hepatic conversion of vitamin D to 25(OH)D is unsettled. Initial studies in vitro showed a marked inhibition of the hepatic enzyme by high levels of the substrate vitamin D, but

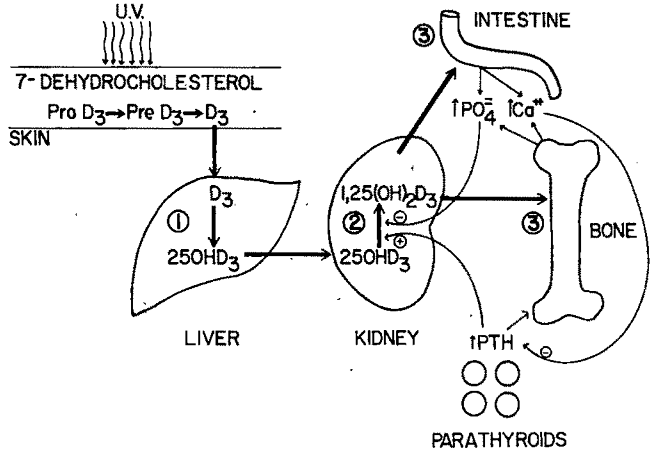


Figure 2 Schema depicting proposed points of regulation in the pathway of vitamin D metabolism.

① 25-Hydroxylation in the liver; ② 1-Hydroxylation in the kidney. Parathyroid hormone stimulates and hyperphosphatemia suppresses the conversion of $25(OH)D_3$ to $1,25(OH)_2D_3$; ③ Action of $1,25(OH)_2D_3$ on cytosolic receptors at intestine and bone, the principal target organs.

subsequent studies contradicted these results (27). In rats (28), blood levels of 25(OH)D_3 rise in proportion to the administered dose of vitamin D_3, and the same linear relationship has also been found in man (27). Most evidence points to the renal-25(OH)D_3-1α-hydroxylase as the principal site of regulation of vitamin D activity, and the recognition of this regulation by phosphate and PTH has provided possible explanations for the existence of some of the vitamin D–resistant osteomalacias.

DEFECTS AT THE LEVEL OF 25-HYDROXYVITAMIN D PRODUCTION

As early as 1968 it became evident that a substantial number of patients receiving chronic anticonvulsant therapy developed symptomatic as well as asymptomatic osteomalacia (29). Hypocalcemia accompanied by elevations of parathyroid hormone and alkaline phosphatase was seen in patients receiving diphenylhydantoin and phenobarbital, either alone or in combination. The severity of osteomalacia was related to their dietary vitamin D intake and degree of sunlight exposure. Serum 25(OH)D levels were found to be low when compared with those of control subjects on similar dietary vitamin D intakes. Hahn et al (30) demonstrated that ^{3}H-labelled vitamin D_3 given to patients on anticonvulsant therapy is converted to 25(OH)D_3, which is then rapidly metabolized to inactive compounds. Formation of similar compounds has also been shown in vitro after incubation, with tritiated vitamin D_3, of hepatic microsomes from diphenylhydantoin-treated animals (30). The resultant decrease in the availability of 25(OH)D substrate would logically cause decreased production of 1,25(OH)$_2$D and development of osteomalacia in these patients. However, recent measurements of serum 1,25(OH)$_2$D levels in patients receiving one or both drugs have been found to be normal (31) despite confirmed low 25(OH)D levels. This observation argues against deficiency of 1,25(OH)$_2$D production as the sole cause underlying anticonvulsant osteomalacia. Whether 25(OH)D itself exerts some influence in mineralization of bone that is independent of 1,25(OH)$_2$D is not known. It is possible that, independent of hepatic effects, anticonvulsants may have effects that occur later in the pathway of vitamin D action, i.e. at the level of the target organs.

Decreased production of 25(OH)D, however, does seem to be the cause of osteomalacia in some patients with severe liver disease. It is known that the 25-hydroxylase is present in the microsomal fraction of normal hepatic tissue; its reactions are independent of cytochrome P450 or lipid-peroxidation mechanisms (32). An enterohepatic circulation of 25(OH)D that requires an intact hepatobiliary system has been demonstrated in man. It is not surprising therefore that alcoholic hepatitis, cirrhosis, and primary biliary cirrhosis all may cause decreased circulating levels of 25(OH)D with a prolonged half-life of vitamin D_3. When nine patients with advanced liver disease were challenged parenterally with 100,000 units of vitamin D_3, all but two failed to show increases in serum 25(OH)D levels (33). Prolonged parenteral therapy for 6 months in 27 patients with primary biliary cirrhosis eventually resulted in elevations of their 25(OH)D levels without, at that time, any healing of their osteomalacia (33). On the other hand, healing of vitamin D–resistant bone disease in primary biliary cirrhosis by administration of oral 25(OH)D was reported (34).

DEFECTS AT THE LEVEL OF 1,25(OH)$_2$D PRODUCTION

Hereditary Vitamin D–Dependent Rickets

Not long after the discovery of vitamin D as the antirachitic factor, it became apparent that there existed a group of rachitic and osteomalacic disorders that remained refractory to treatment by vitamin D in amounts sufficient to cure nutritional vitamin D–deficiency rickets (35). Many of these vitamin D–resistant forms of the disease were clearly hereditary. One such disorder is hereditary vitamin D–dependent rickets or "pseudovitamin D–deficiency" rickets (36). The disease is inherited in an autosomal recessive fashion and differs from other vitamin D–resistant disorders in that treatment with large doses of vitamin D brings about a complete remission of the disease. The elucidation of the metabolic pathway for vitamin D raised the possibility that one or more of these disorders might be due to a defect in the metabolism of the vitamin. Fraser et al (37) reported evidence, albeit indirect, indicating that hereditary vitamin D–dependent rickets results from a block in the conversion of 25(OH)D$_3$ to 1,25(OH)$_2$D$_3$, and they suggested an inherited deficiency in the 25(OH)D-1α hydroxylase. In five children with the disorder, large doses of vitamin D$_3$, on the order of 1.25 mg (50,000 units) per day, were required to heal rachitic lesions. However, by treatment with 1 μg of 1,25(OH)$_2$D$_3$ per day, complete healing was achieved. These children were also refractory to 25(OH)D$_3$ and required 200–900 μg per day for effective treatment. Thus, the ratios of the dose requirements of vitamin D:25(OH)D$_3$:1,25(OH)$_2$D$_3$ were on the average 1700:700:1. Inasmuch as the dose-requirement ratio for vitamin D$_3$:25(OH)D$_3$ (3:1) observed in these studies is similar to that found in rachitic rats (38) and in two infants with vitamin D deficiency (39), the findings point to the existence of a specific block at the level of the conversion of 25(OH)D to 1,25(OH)$_2$D and not at the level of conversion of D to 25(OH)D. Recently, these suppositions have been strengthened by direct measurements of 1,25(OH)$_2$D in blood of the children; the levels were low and returned to normal during treatment with 1.2–1.8 μg per day of 1α(OH)D$_3$ (40).

Hypoparathyroidism and Pseudohypoparathyroidism

Although hypoparathyroidism and pseudohypoparathyroidism do not usually present with osteomalacia or metabolic bone disease (40a,b) there is evidence that in some patients with these diseases one of the biochemical aberrations is a block in the synthesis of 1,25(OH)$_2$D from 25(OH)D. This abnormality may explain why patients with these disorders require massive doses of vitamin D (up to 200–300 times the normal requirement) to prevent hypocalcemia. Several reports have established that $1,25(OH)_2D_3$ or $1\alpha(OH)D_3$ is approximately 100 times more effective than vitamin D_3 on a weight basis in controlling the biochemical abnormalities in these disorders (39, 41–43). In addition, circulating levels of 25(OH)D are appropriately elevated in patients treated with high doses of 25(OH)D, but no improvement in hypocalcemia is observed. Blood levels of $1,25(OH)_2D$ measured by competitive-protein-binding assays have been reported to be low in both disorders (40, 42, 43).

The finding of a defect in the conversion of 25(OH)D to $1,25(OH)_2D$ in the syndrome of pseudohypoparathyroidism suggests that inhibitory effects of hyperphosphatemia may override the stimulatory effects of PTH on the 25(OH)D-1α-hydroxylase. Patients with pseudohypoparathyroidism are known to have elevated blood levels of PTH; the basic defect in this disorder is believed to be an end-organ resistance to the actions of PTH (44). It remains possible, however, that the syndrome of pseudohypoparathyroidism may include a refractoriness of the renal 1α-hydroxylase to the action of PTH.

One puzzling aspect of these diseases is the apparent absence of overt metabolic bone disease. Although short stature often accompanies both disorders when present in children, rickets or osteomalacia is seen only occasionally (40a, b). Likewise, patients with pseudohypoparathyroidism in whom PTH levels in blood are elevated, usually do not develop osteitis fibrosa or subperiosteal resorption.

Nephrectomy and/or Severe Renal Failure

It has been appreciated for some time that patients with severe renal disease may develop bone disease with an osteomalacic component that is resistant to treatment with normal doses of vitamin D. Several lines of evidence now support the concept that in some, if not all, of these patients a deficiency in the ability of the failing kidney to produce $1,25(OH)_2D$ contributes to the development of the osteomalacia. First, it has been shown that plasma levels of 25(OH)D are normal in most patients with renal failure (45). Thus, vitamin D deficiency per se is not a problem in these patients as long as normal dietary intake of vitamin D is maintained and the patient does not have intestinal malabsorption. Second, direct measurements of $1,25(OH)_2D$ levels in blood of patients with renal failure reveal values that are lower than normal, and patients who are anephric have undetectable levels as assayed by current methods (5, 20). Finally, impressive healing of osteomalacia and osteitis fibrosa has been observed in many renal patients treated with near-physiologic doses of synthetic $1,25(OH)_2D_3$ (46).

This cumulative information suggests that resistance of target organs to the actions of $1,25(OH)_2D$ is not as important as a deficiency in the production of $1,25(OH)_2D$. Moreover, this deficiency of $1,25(OH)_2D$, by way of chronic hypocalcemia may aggravate secondary hyperparathyroidism, the other component of renal osteodystrophy.

Whether $1,25(OH)_2D$ deficiency may be reversible in some patients is also a matter of considerable importance. Chronically elevated serum phosphate may be, in part, responsible for the inhibition of the formation of $1,25(OH)_2D$. If this should prove to be true, then control of hyperphosphatemia by itself may return $1,25(OH)_2D$ levels to normal in those patients who have sufficient functioning renal tissue.

As yet, no information is available that could be used to predict which renal patients will develop osteomalacia, or at what point in their disease this will occur. In fact, evidence suggests that most patients with renal disease do not develop osteomalacia. For example, a recent study by Kanis et al (47) indicated that only 16% of 93 patients studied before the beginning of treatment by hemodialysis had evidence of osteomalacia by bone biopsy. Furthermore, when restudied prospectively, the incidence of osteomalacia did not increase, even in those who underwent nephrectomy. Why some patients develop osteomalacia and others do not remains an enigma; no differences in the multiple factors involved in their mineral and skeletal metabolism have been identified. Studies of $1,25(OH)_2D$ production rates may eventually answer some of these questions.

Oncogenic Osteomalacia

Over the years there have been several reports of patients with severe osteomalacia in association with giant-cell tumors of bone or mesenchymal tumors (48). After removal of the tumors the osteomalacia syndrome subsided. Recently Drezner & Feinglos (49), reported a patient with oncogenic osteomalacia and a giant-cell tumor of bone in whom blood levels of $25(OH)D$ were normal but those of $1,25(OH)_2D$ were reduced. Treatment of the patient with $1,25(OH)_2D_3$ resulted in normalization of serum phosphorus, renal-tubular maximum for reabsorption of phosphate, aminoaciduria, intestinal calcium and phosphorus absorption, and remineralization of bone. The causal role of the tumor in the development of the osteomalacia syndrome was confirmed by finding that, after total resection of the tumor, all parameters of calcium and phosphorus metabolism measured returned to normal. The evidence suggests that certain tumors of connective-tissue origin produce a substance that inhibits the conversion of $25(OH)D$ to $1,25(OH)_2D$. It is noteworthy that the syndrome of oncogenic osteomalacia resembles in virtually all respects the osteomalacias associated with primary renal-phosphate wasting, i.e. hereditary vitamin D–resistant osteomalacias (known as phosphate diabetes). In this disorder not associated with neoplasms, however, therapy with either $25(OH)D_3$ (50) or $1,25(OH)_2D_3$ (51) has thus far been ineffective. Moreover, it is likely that in phosphate diabetes primary defects in the renal reabsorption and intestinal absorption of phosphate, and not defects in vitamin D metabolism, are responsible for the hypophosphatemia and osteomalacia (52, 53).

IMPAIRMENT OF ACTION OF $1,25(OH)_2D$ ON TARGET TISSUES

New information suggests that situations may exist in which adequate amounts of active vitamin D metabolite, $1,25(OH)_2D$, are produced, but the target tissues of intestinal mucosa, bone, and kidney may not respond normally. An example of this unresponsiveness of target tissue may exist in anticonvulsant osteomalacia. As discussed previously, the recent finding of normal circulating levels of $1,25(OH)_2D$ in patients receiving diphenylhydantoin and/or phenobarbital has weakened previous theories placing the defect in anticonvulsant osteomalacia solely in the shunting of 25(OH)D substrate away from production of $1,25(OH)_2D$. It is possible that anticonvulsant drugs induce target-organ resistance to $1,25(OH)_2D$, either at the intestinal level (decreased calcium absorption) or at the bone level (inhibition of normal mineralization), or both.

Several studies indicate that diphenylhydantoin and phenobarbital directly inhibit intestinal calcium absorption in rats (54) as well as calcium transport in chicken duodena in vitro (55). Most recently, the work of Harrison & Harrison (56) demonstrated that combined diphenylhydantoin and phenobarbital therapy causes an inhibition of energy-dependent calcium transport in intestinal mucosa of vitamin D–depleted animals when challenged with vitamin D_2, vitamin D_3, or $25(OH)D_3$. These findings are consistent with the course of a patient with hypoparathyroidism who required phenobarbital while being treated with $1,25(OH)_2D$. This patient required five times more $1,25(OH)_2D$ to control her serum calcium than did hypoparathyroid patients not taking anticonvulsants (43).

In addition to its effects on the intestine, there is evidence that diphenylhydantoin may alter the responsiveness of bone to vitamin D metabolites. Studies in organ cultures of mouse calvaria demonstrated that therapeutic levels of the drug inhibit bone resorption that is normally stimulated by 25(OH)D as well as by PTH (57). This may result from diphenylhydantoin-induced alterations in membrane permeability to sodium ions, which indirectly diminish calcium influx, similar to the role proposed for sodium ions in the cellular transport of calcium in both intestine and bone.

Another example of resistance of target tissue to vitamin D is seen in the decreased calcium absorption found in uremia. Multiple studies confirmed that this reduction in calcium absorption varies with the degree of uremia. Specifically, there appears to be a defect in calcium transport in the proximal part of the small intestine, where vitamin D may be most active in stimulating calcium absorption (58). In light of the work of Brickman et al (59), this defect does not appear to be completely due to a deficiency of $1,25(OH)_2D$. In their study, normal subjects responded to $1,25(OH)_2D_3$ doses of 0.14–2.7 μg/day with clear increases in calcium absorption, sometimes to levels above the normal range. However, doses of $1,25(OH)_2D_3$ several-fold higher than those used in normal subjects were required to elicit a response in uremic patients. This finding is compatible with data in experimental animals with uremia and suggests that uremia per se contributes to impaired intestinal absorption of calcium through some as-yet undefined effect on intestinal mucosa.

One area of much controversy in recent years has been that of steroid-induced bone disease. Although steroids do not produce osteomalacia, in pharmacologic doses they do decrease intestinal calcium absorption and with prolonged therapy result in osteoporosis. The precise mechanism by which this reduction in calcium absorption is initiated remains unknown. Earlier investigations in this area suggested that corticosteroids administered over a long time altered vitamin D metabolism in man (60). Subsequently, in vitro studies of rats failed to demonstrate any alterations in hepatic or renal hydroxylation, and therefore suggest that cortiocosteroids diminish calcium absorption through disruption of the calcium transport process in a manner unrelated to vitamin D (61). Currently, studies of patients receiving trials of therapy with corticosteroids in high and low doses, either every day or on alternate days, have revealed that a decreased calcium absorption occurs only with high doses of prednisone (>15 mg/day) given daily. The decrease was not seen in patients receiving either alternate-day steroid treatment or low, near-physiologic daily doses of $1,25(OH)_2D_3$ ($0.4\ \mu$g/day) (62). Resolution of these conflicting data must await better techniques for the analyses of actual $1,25(OH)_2D_3$ production rates and the physiologic effects of $1,25(OH)_2D_3$ on target organs.

THERAPEUTIC CONSIDERATIONS

In recent years, clinical trials have been initiated to test the therapeutic efficacy of $25(OH)D_3$, $1,25(OH)_2D_3$, and analogues of vitamin D_3 such as $1\alpha(OH)D_3$ and 5,6-trans-vitamin D (Figure 1B) in patients who have vitamin D–resistant forms of osteomalacia and in other patients with presumed disorders in the metabolism of vitamin D (hypoparathyroidism and pseudohypoparathyroidism). These compounds have been prepared synthetically and are available for clinical research studies. The rationale behind the use of these compounds is that they are hydroxylated at either or both of the 1 and 25 positions, and, as such, bypass metabolic blocks that might exist at the level of the hydroxylation reactions in the liver and the kidney.

An additional advantage in the use of these hydroxylated compounds is that the time required for their onset of action and for their clearance from the body is much faster than it is for vitamin D. For example, in a comparative study of the relative effectiveness of vitamin D_2, dihydrotachysterol, $25(OH)D_3$, and $1,25(OH)_2D_3$ in the treatment of patients with hypoparathyroidism, it was found that the time required to restore serum calcium to normal was 4–8 wk for vitamin D_3, 1–2 wk for dihydrotachysterol, 2–4 wk for $25(OH)D_3$, and 0.5–1 wk for $1,25(OH)_2D_3$ (63). Likewise, toxic effects, when they occur by inadvertent overdosage with these compounds, subside more rapidly upon cessation of therapy with the hydroxylated compounds than with vitamin D: 6–18 wk for vitamin D_3, 1–3 wk for dihydrotachysterol, 4–12 wk for $25(OH)D_3$, and 0.5–1 wk for $1,25(OH)_2D_3$.

The use of $25(OH)D_3$ in the treatment of patients with osteomalacia and liver disease appears to hold some promise (64). Whether it will prove to have advantages over treatment with larger doses of the parent compound, vitamin D_3, remains to be determined.

Good results have been obtained in patients with anticonvulsant osteomalacia who were treated with $25(OH)D_3$ at dosage levels of $40\ \mu$g/day (65, 66). Blood levels of calcium, phosphorus, alkaline phosphatase, and PTH all returned to normal. Similar good results, however, have been obtained by treatment with low doses of vitamin D_3 (3000–10,000 u/wk) (66–68). It has not been established whether $25(OH)D_3$ has any significant therapeutic advantages over vitamin D_3 in the treatment of this disorder.

The use of 1-hydroxylated vitamin D compounds in patients with defects in renal 1-hydroxylation appears to hold more promise than does the use of $25(OH)D_3$ in patients with presumed defects in hepatic 25-hydroxylation. Control of hypocalcemia and hyperphosphatemia in patients with hereditary vitamin D–dependent rickets has been achieved by treatment with near-physiologic doses of $1,25(OH)_2D_3$ (mean dose of 1.4 μg/day). Similarly low doses of $1\alpha(OH)D_3$ have been adequate to effectively treat hypoparathyroidism (63) and pseudohypoparathyroidism (42). In a study of the relative effectiveness of the different vitamin D compounds in treatment of hypoparathyroidism, equivalent responses were obtained with doses of 0.75–3.0 mg of vitamin D_3, 0.25–1.0 mg of dihydrotachysterol, 50–200 μg of $25(OH)D_3$, and 0.5–2.0 μg of $1,25(OH)_2D_3$ (63).

The treatment of pseudohypoparathyroidism with $1,25(OH)_2D_3$ and $1\alpha(OH)D_3$ has been reported by several groups (39, 42). The effective doses ranged from 0.5–5.0 μg/day. Of particular therapeutic interest is the suggestion by Brickman et al (42) that patients with pseudohypoparathyroidism may be a more heterogeneous group with respect to their response to 1-hydroxylated vitamin D analogues than are hypoparathyroid patients. In over 20 patients studied they noted that mean doses of 1–3 μg/day of $1,25(OH)_2D_3$ were required to achieve normocalcemia (a dose generally 2- to 3-fold that necessary in hypoparathyroidism). However, after several months of therapy, the drug could be decreased to 25% of the initial dose. Presumably, this reduction in dosage is a result of improvement in the underlying metabolic bone disease that exists in some of these patients. The possibility should be kept in mind that when patients are treated with the potent hydroxylated forms of vitamin D adjustments in dosage levels may be required to avoid toxicity. It should also be emphasized that it is important to provide the patient with adequate amounts of dietary and/or supplemental calcium and phosphate during therapy.

By far the most heterogeneous group of patients that have been treated with vitamin D compounds is that with chronic renal failure. Multiple factors influence their status with respect to bone disease and vitamin D metabolism. Such factors include age, duration of renal failure, prior treatment, degree of secondary hyperparathyroidism or osteomalacia, and the types of treatment used concurrently with the vitamin D compounds (hemodialysis, corticosteroids, anticonvulsant drugs, dietary restrictions). The variability of these factors as well as the relative degrees of osteomalacia and secondary hyperparathyroidism render it difficult to conduct and interpret large-scale trials of the efficacy of new vitamin D compounds. Treatment successes, treatment failures, and toxicity secondary to therapy have been reported with virtually all the analogues used to date: $1\alpha(OH)D$, $1,25(OH)_2D$, $25(OH)D$, dihydrotachysterol

(DHT), 5,6-trans-vitamin D_3 (69–71). Some studies indicated that treatment of patients with severe renal failure with $1,25(OH)_2D_3$ can lead to remarkable improvement in both the symptoms and the radiographic appearance of osteomalacia (46).

At the present time one can only state that the use of 1-hydroxylated forms of vitamin D holds considerable promise for the treatment of the disabling osteomalacia that may accompany renal failure. More experience in the use of these 1-hydroxylated vitamin D compounds must be gathered before appropriate therapy can be chosen for the individual patient with renal osteodystrophy.

CONCLUSION

A rapidly growing understanding of the biochemical and physiological processes that underlie the metabolism of vitamin D has provided new insights into the pathogenesis of osteomalacia. Many of the vitamin D–resistant osteomalacia syndromes can now be explained on the basis of defects in the metabolic conversion of vitamin D to the biologically active dihydroxylated metabolite $1,25(OH)_2D$ and perhaps, in some instances, to impairment of the actions of $1,25(OH)_2D$ on target tissues. The availability of this new information has made possible the synthesis of 1-hydroxylated forms of the vitamin for therapeutic use in states of vitamin D resistance. Although many questions regarding the pathogenesis and most effective approaches in the management of osteomalacia remain unanswered, considerable progress has been made in this direction as a result of continued research on the subject.

ACKNOWLEDGMENTS

We thank M. F. Holick, M. B. Clark, and L. B. Fred for helpful suggestions in the preparation of this chapter.

Literature Cited

1. Arnstein, A. R., Frame, B., Frost, H. M. 1967. Recent progress in osteomalacia and rickets. *Ann. Intern. Med.* 67:1296–1330
2. Stanbury, S. W. 1972. Osteomalacia. *Clinics Endocrinol. Metab.* 1:239–67
3. Mankin, H. J. 1974. Rickets, osteomalacia and renal osteodystrophy. *J. Bone Jt. Surg.* 56A:101–28
4. DeLuca, H. F. 1973. The kidney as an endocrine organ for the production of 1,25-dihydroxyvitamin D_3. *N. Engl. J. Med.* 289:359–65
5. Haussler, M. R., Baylink, D. J., Hughes, M. R., Brumbaugh, P. F., Wergedal, J. E., Shen, F. H., Nielsen, R. L., Counts, S. J., Bursac, K. M., McCain, T. A. 1976. The assay of 1α,25-dihydroxyvitamin D_3: physiologic and pathologic modulation of circulating hormone levels. *Clin. Endocrinol.* 5:151s–65s
6. Norman, A. W., Henry, H. 1974. The role of the kidney and vitamin D metabolism in health and disease. *Clin. Orthop. Relat. Res.* 98:258–87
7. DeLuca, H. F., Schnoes, H. K. 1976. Metabolism and mechanism of action of vitamin D. *Ann. Rev. Biochem.* 45:631–66
8. Wasserman, R. H., Corradino, R. A., Fullmer, C. S., Taylor, A. N. 1974. Some aspects of vitamin D action; calcium absorption and vitamin D-dependent calcium-binding protein. *Vitam. Horm. NY* 32:299–324

9. Pommer, G. 1885. *Untersuchungen über Osteomalacia und Rachitis.* Leipzig: FCW Voegel. 514 pp.
10. Hess, A. F., Unger, L. J. 1917. Prophylactic therapy for rickets in a Negro community. *J. Am. Med. Assoc.* 69:1583–86
11. Huldschinsky, K. 1919. Heilung von Rachitis durch kunstliche Hohensonne. *Dtsch. Med. Wochenschr.* 45:712–13
12. McCollum, E. V., Simmonds, N., Becker, J. E., Shipley, P. G. 1922. Studies in experimental rickets. XXI. An experimental demonstration of the existence of a vitamin which promotes calcium deposition. *J. Biol. Chem.* 53:293–312
13. Windaus, A. 1931. Chemistry of irradiated ergosterol. *Proc. R. Soc. Med.* 108:568–70
14. Lund, J., DeLuca, H. F. 1966. Biologically active metabolite of vitamin D_3 from bone, liver and blood serum. *J. Lipid Res.* 7:739–44
15. Ponchon, G., DeLuca, H. F. 1969. The role of the liver in the metabolism of vitamin D. *J. Clin. Invest.* 48:1273–79
16. Lawson, D. E. M., Fraser, D. R., Kodicek, E., Morris, H. R., Williams, D. H. 1971. Identification of 1,25-dihydroxycholecalciferol, a new kidney hormone controlling calcium metabolism. *Nature* 230:228–30
17. Holick, M. F., Schnoes, H. K., DeLuca, H. F., Suda, T., Cousins, R. J. 1971. Isolation and identification of 1,25-dihydroxycholecalciferol: a metabolite of vitamin D active in intestine. *Biochemistry* 10:2799–2804
18. Haddad, J. G., Chyu, K. J. 1971. Competitive protein-binding radioassay for 25-hydroxycholecalciferol. *J. Clin. Endocrinol. Metab.* 33:992–95
19. Eisman, J. A., Hamstra, A. J., Kream, B. E., DeLuca, H. F. 1976. 1,25-Dihydroxyvitamin D in biological fluids: a simplified and sensitive assay. *Science* 193:1021–23
20. Brumbaugh, P. F., Haussler, D. H., Bressler, R., Haussler, M. R. 1974. Radioreceptor assay for 1α,25-dihydroxyvitamin D_3. *Science* 183:1089–91
21. Boyle, I. T., Gray, R. W., DeLuca, H. F. 1971. Regulation by calcium of *in vivo* synthesis of 1,25-dihydroxycholecalciferol and 21,25-dihydroxycholecalciferol. *Proc. Natl. Acad. Sci. USA.* 68:2131–34
22. Garabedian, M., Holick, M. F., Deluca, H. F., Boyle, I. T. 1972. Control of 25-hydroxycholecalciferol metabolism by parathyroid glands. *Proc. Natl. Acad. Sci. USA* 69:1673–76
23. Fraser, D. R., Kodicek, E. 1973. Regulation of 25-hydroxycholecalciferol-1-hydroxylase activity in kidney by parathyroid hormone. *Nature* 241:163–66
24. Rasmussen, H., Wong, M., Bikle, D , Goodman, D. B. P. 1972. Hormonal control of the renal conversion of 25-hydroxycholecalciferol to 1,25-dihydroxycholecalciferol. *J. Clin. Invest.* 51:2502–4
25. Tanaka, Y., Frank, H., DeLuca, H. F. 1973. Intestinal calcium transport: stimulation by low phosphorus diets. *Science* 181:564–66
26. Tanaka, Y., DeLuca, H. F. 1973. The control of 25-hydroxyvitamin D metabolism by inorganic phosphorus. *Arch. Biochem. Biophys.* 154:566–74
27. Bhattacharyya, M. H., DeLuca, H. F. 1974. The regulation of calciferol 25-hydroxylase in the chick. *Biochem. Biophys. Res. Commun.* 59:734–41
28. Clark, M. B., Potts, J. T. Jr. 1976. 25-Hydroxyvitamin D_3 regulation. *Calcif. Tissue Res.* 22:Suppl. 29–34
29. Kruse, R. 1968. Osteopathien bei antiepileptischer Langzeittherapie. *Monatsschr. Kinderheilkd.* 116:378–81
30. Hahn, T. J., Birge, S. J., Scharp, C. R., Avioli, L. V. 1972. Phenobarbital-induced alterations in vitamin D metabolism. *J. Clin. Invest.* 51:741–48
31. Jubiz, W., Haussler, M. R., McCain, T. A., Tolman, K. G. 1977. Plasma 1,25-dihydroxyvitamin D levels in patients receiving anticonvulsant drugs. *J. Clin. Endocrinol. Metab.* 44:617–21
32. Delvin, E. E., Bourbonnais, R., Glorieux, F. H. 1977. Effect of phenobarbital on hepatic vitamin D-25-hydroxylase and mixed function oxydases activities in vitro. In *Vitamin D: Biochemical, Chemical and Clinical Aspects Related to Calcium Metabolism: Proc. Workshop Vitamin D, 3rd, Asilomar, California, 1977*, ed. A. Norman. K. Shaefer, J. Coburn, H. DeLuca, D. Fraser,

H. Grigoleit, D. von Herrath, pp. 767–69. New York: de Gruyter. 973 pp.

33. Skinner, R. K., Long, R. G., Sherlock, S., Wills, M. R. 1977. 25-Hydroxyvitamin D values in chronic liver disease and response to treatment with parenteral vitamin D. See Ref. 32, pp. 839–41

34. Hahn, T. J., Avioli, L. V. 1977. Hepatic bioactivation of vitamin D and clinical implications. See Ref. 32, pp. 737–41

35. Albright, F., Butler, A. M., Bloomberg, E. 1937. Rickets resistant to vitamin D therapy. *Am. J. Dis. Child.* 54:529–47

36. Scriver, C. R. 1970. Vitamin D dependency. *Pediatrics* 45:361–63

37. Fraser, D., Kooh, S. W., Kind, H. P. 1973. Pathogenesis of hereditary vitamin-D-dependent rickets. *N. Engl. J. Med.* 289:817–22

38. Tanaka, Y., Frank, H., DeLuca, H. F. 1973. Biological activity of 1,25-dihydroxycholecalciferol in the rat. *Endocrinology* 92:417–22

39. Kooh, S. W., Fraser, D., DeLuca, H. F., Holick, M. F., Belsey, R. E., Clark, M. B., Murray, T. M. 1973. Treatment of hypoparathyroidism and pseudohypoparathyroidism with metabolites of vitamin D: evidence for impaired conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D. *N. Engl. J. Med.* 289:817–22

40. Frame, B., Parfitt, A. M. 1977. Vitamin D dependent bone disease: long-term response to vitamin D analogs and effect on tetracycline based bone dynamics. See Ref. 32, pp. 373–75

40a. Drezner, M. K., Neelon, F. A., Jowsey, J., Lebovitz, H. E. 1977. Hypoparathyroidism: a possible cause of osteomalacia. *J. Clin. Endocrinol. Metab.* 45:114–22

40b. Albright, F. 1956. Hypoparathyroidism as a cause of osteomalacia. *J. Clin. Endocrinol. Metab.* 16:419–25

41. Neer, R. M., Holick, M. F., DeLuca, H. F., Potts, J. T. Jr. 1975. Effects of 1α-hydroxy-vitamin D_3 and 1,25-dihydroxy-vitamin D_3 on calcium and phosphorus metabolism in hypoparathyroidism. *Metabolism* 24:1403–13

42. Brickman, A. S., Norman, A. W., Coburn, J. W. 1977. Vitamin D and pseudo hypoparathyroidism. See Ref. 32, pp. 867–76

43. Rosen, J. F., Fleishman, A. R. Finberg, L., Eisman, J., DeLuca, H. F. 1977. 1,25-Dihydroxycholecalciferol: oral administration and sterol levels in the long-term management of idiopathic hypoparathyroidism in children. See Ref. 32, pp. 827–30

44. Potts, J. T. Jr. 1978. Pseudohypoparathyroidism. In *The Metabolic Basis of Inherited Disease*, 4th edition, ed. J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson. New York: McGraw-Hill.

45. Eastwood, J. B., Harris, E., Stamp, T. C., DeWordener, H. E. 1976. Vitamin-D deficiency in the osteomalacia of chronic renal failure. *Lancet* 2:1209–11

46. Coburn, J. W., Brickman, A. S., Sherrard, D. J., Singer, F. R., Baylink, D. J., Wong, E. G. C., Massry, S. G., Norman, A. W. 1977. Clinical efficacy of 1,25-dihydroxyvitamin D_3 in renal osteodystrophy. See Ref. 32, pp. 657–66

47. Kanis, J. A., Adams, N. D., Earnshaw, M., Heynen, G., Ledingham, J. G. G., Oliver, D. O., Russell, R. G. G., Woods, C. G. 1977. Vitamin D, osteomalacia, and chronic renal failure. See Ref. 32, pp. 671–73

48. Salasso, R. M., Jowsey, J., Arnaud, C. D. 1970. Hypophosphatemic osteomalacia associated with "non-endocrine" tumors. *N. Engl. J. Med.* 283:65–70

49. Drezner, M. K., and Feinglos, M. N. 1977. Tumor induced 1α-25 dihydroxycholecalciferol (1α,25-$(OH)_2D_3$) deficiency: a cause of oncogenic osteomalacia. See Ref. 32, p. 863

50. Earp, H. S., Ney, R. L., Gitelman, H. J., Richman, R., DeLuca, H. F. 1970. Effects of 25-hydroxycholecalciferol in patients with familial hypophosphatemia and vitamin-D-resistant rickets. *N. Engl. J. Med.* 283:627–30

51. Brickman, A. S., Coburn, J. W., Kurokawa, K. Bethune, J. E., Harrison, H. E., Norman, A. W. 1973. Actions of 1,25-dihydroxycholecalciferol in patients with hypophosphatemic, vitamin-D-resistant rickets. *N. Engl. J. Med.* 289:495–98

52. Glorieux, F., Scriver, C. R. 1972.

Loss of a parathyroid hormone–sensitive component of phosphate transport in X-linked hypophosphatemia. *Science* 175:997–1000

53. Short, E. M., Binder, H. J., Rosenberg, L. E. 1973. Familial hypophosphatemic rickets: defective transport of inorganic phosphate by intestinal mucosa. *Science* 179:700–2
54. Caspary, W. F. 1972. Inhibition of intestinal calcium transport by diphenylhydantoin in rat duodenum. *Naunyn-Schmiedebergs Arch. Pharmakol.* 274:146–53
55. Corradino, R. A. 1976. Diphenylhydantoin: direct inhibition of vitamin D_3-mediated calcium absorptive mechanism in organ cultured duodenum. *Biochem. Pharmacol.* 25:863–64
56. Harrison, H. C., Harrison, H. E. 1976. Inhibition of vitamin D-stimulated active transport of calcium of rat intestine by diphenylhydantoin-phenobarbital treatment. *Proc. Soc. Exp. Biol. Med.* 153:220–24
57. Jenkins, M. V., Harris, M., Wills, M. R. 1974. The effect of phenytoin on parathyroid extract and 25-hydroxycholecalciferol-induced bone resorption: adenosine 3,'5' cyclic monophosphate production. *Calcif. Tissue Res.* 16:163–67
58. Brickman, A. S., Massry, S. G., Norman, A. W., Coburn, J. W. 1975. On the mechanism and nature of the defect in intestinal absorption of calcium in uremia. *Kidney Int. Suppl.:* S113–S117
59. Brickman, A. S., Coburn, J. W., Friedman, G. A., Okamura, W. H., Massry, S. G., Norman, A. W. 1976. Comparison of effects of 1α-hydroxy-vitamin D_3 and 1,25-dihydroxy-vitamin D_3 in man. *J. Clin. Invest.* 57:1540–47
60. Avioli, L. V., Birge, S. J., Lee, S. W. 1968. Effects of prednisone on vitamin D metabolism in man. *J. Clin. Endocrinol.* 28:1341–46
61. Kimberg, D. V., Baerg, R. D., Gershon, E., Graudusius, R. T. 1971. Effect of cortisone treatment on the active transport of calcium by the small intestine. *J. Clin. Invest.* 50:1309–21
62. Klein, R. G., Arnaud, S. B., Gallagher, J. L., DeLuca, H. F., Riggs, B. L. 1977. Intestinal calcium absorption in exogenous hypercortisonism. *J. Clin. Invest.* 60:253–59
63. Parfitt, A. M. 1977. The treatment of hypoparathyroidism with 25(OH) D_3: practical conclusions and metabolic implications. See Ref. 32, pp. 809–18
64. Wagonfeld, J. B., Nemchausky, B A., Bolt, M., VanderHorst, J., Boyer, J. L., Rosenberg, I. H. 1976. Comparison of vitamin D and 25-hydroxy-vitamin D in the therapy of primary biliary cirrhosis. *Lancet* 2:391–94
65. Stamp, T. C. B., Round, S. M., Rowe, D. J. F., Haddad, J. G. 1972. Plasma levels and therapeutic effect of 25-hydroxycholecalciferol in epileptic patients taking anticonvulsant drugs. *Br. Med. J.* 4:9–12
66. Hahn, J. T. 1976. Bone complications of anticonvulsants. *Drugs* 12:201–11
67. Christiansen, C., Rodbro, P., Lund, M. 1973. Effect of vitamin D on bone mineral mass in normal subjects and in epileptic patients on anticonvulsants: a controlled therapeutic trial. *Br. Med. J.* 2:208–9
68. Offermann, G., Kruse, R. 1977. Prophylactic vitamin D supplementation in chronic anticonvulsant therapy. See Ref. 32, pp. 805–7
69. Cordy, P. E. 1976. Treatment of bone disease in patients on chronic hemodialysis with dihydrotachysterol. *Trans. Am. Soc. Artif. Intern. Organs* 22:60–63
70. Schulz, W., Heidler, R., Spiegel, P., Gessler, U., Schulz, A., Delling, G. 1977. The treatment of renal osteopathy with 5,6-trans-25OHCC. A report of experience made during 32 months therapy. See Ref. 32, pp. 715–18
71. Recker, R. R., Schoenfeld, P., Letteri, J., Slatopolsky, E., Martin, K., Kleinman, L., Hartenbower, D., Brickman, A., Goldsmith, R., Frost, H., Lee, W., Teitelbaum, S., Kimmel, D., Arnaud, C., Arnaud, S., Heaney, R. 1977. 25 Hydroxyvitamin D in renal osteodystrophy: results of a six-center trial: Preliminary report. See Ref. 32, pp. 649–56
72. Holick, M. F., Frommer, J. E., McNeill, S. C., Richtand, N. M., Henley, J. W., Potts, J. T. Jr. 1977. Photometabolism of 7-dehydrocholesterol to previtamin D_3 in skin. *Biochem. Biophys. Res. Commun.* 76:107–14

Ann. Rev. Med. 1978. 29:343–58
Copyright © 1978 by Annual Reviews Inc. All rights reserved

DIALYSIS VERSUS TRANSPLANTATION IN THE TREATMENT OF END-STAGE RENAL DISEASE

◆7284

John P. Merrill, M.D.
Harvard Medical School, and Peter Bent Brigham Hospital, Boston,
Massachusetts 02115

INTRODUCTION

The title of this chapter is perhaps somewhat misleading since the major question is not dialysis *versus* transplantation but that of combining dialysis and transplantation in the optimum management of the patient with end-state renal disease (ESRD). Indeed, it was the availability of an artificial kidney for hemodialysis that made the first transplants possible since it would obviously have been impossible to await the availability of a cadaver donor in a terminally ill patient with renal failure without some means of keeping him alive and in condition to tolerate surgery (1). There are, of course, some differences between the two modalities, which will be stressed. These include the many metabolic defects present in patients with kidney failure treated with hemodialysis that are correctable by successful transplantation. There are many complications in the management of patients by either technique and some of the complications are common to both. However, there can be no question but that successful transplantation of the kidney is preferable in an identical twin in whom severe bone disease or cardiovascular disease, common in prolonged dialysis, has not occurred. Where there is no recurrence of his original disease, such a patient can live a completely normal life without any detectable physical or metabolic defect. This cannot be said of even the best techniques of chronic dialysis.

CHRONIC DIALYSIS FOR ESRD

There are two forms of chronic dialysis available at present, peritoneal dialysis and hemodialysis. It is now possible to insert catheters into the peritoneal cavity for long periods of time, infuse a rinsing fluid of the proper composition, and

0066-4219/78/0401-0343$01.00

after a suitable "dwell time" remove it and repeat the cycle. Solute diffuses from the blood into the peritoneal fluid and is removed during the exchange. Automated apparatus has been devised so that this process may be carried on continuously in the patient's home, even while he sleeps. Patients treated by this method constitute the smallest portion of all patients on chronic dialysis; but the popularity of the technique is increasing and it is particularly useful in some patients in whom the impossibility of proper access to the circulation makes hemodialysis difficult or impossible (2).

In hemodialysis blood flows on one side of a semipermeable membrane (usually a derivative of cellophane) and dialysis fluid (consisting of normal concentrations of electrolytes and glucose) on the other. Because the concentration of such substances as creatinine and urea are high in the blood of ESRD patients and are not contained in the dialysis fluid, these substances diffuse from the blood into the dialysate where they are removed. In addition, some types of apparatus are devised primarily for hemofiltration, a process in which a pressure gradient is created across the membrane by one of a number of techniques and protein-free fluid (essentially extracellular fluid) is removed. This technique is useful in removing fluid from patients who are overhydrated.

From the early crude rotating drum type of apparatus, many types of devices have evolved. These include the twin coil dialyzer in which the blood flows through a series of cellophane tubes, the flat plate dialyzer in which blood flows between two sheets of membrane outside of which the dialysate passes, and the so-called hollow fiber kidney composed of thousands of small fibers packed in a cylinder. By means of a special "header," blood is circulated through the semipermeable fibers and the dialysis fluid between them, thus effecting the removal of solute (Figure 1). The addition of certain sorbents such as exchange resins and activated charcoal to the bath traps some solutes as they are removed, thus perpetuating the diffusion gradient from blood to bath. More recent devices include the portable and wearable artificial kidney. All the constituents of the portable dialyzer have been miniaturized and can be carried in a case the size of a large briefcase. The dialyzing fluid can be constituted by the addition of the proper electrolyte concentrates to tap water in a hotel and the patient may travel with such an apparatus, dialyzing himself at appropriate intervals (3). The wearable artificial kidney also makes use of sorbents to decrease the volume of dialysis fluid. It weighs about 8 lb and is strapped around the patient's chest. The wearer can move around but at the present time has to connect the dialyzing fluid circulation to a 20 l tank at least half the dialysis time and usually can disconnect himself only for 15 minutes at a time (4). The device is not a practical therapy at the present time but perhaps with the emergence of better sorbents it will be.

Chronic dialysis over long periods of time was first made possible by the introduction of the arteriovenous shunt by Quinton & Scribner (5). Teflon-silastic catheters were inserted into the radial artery and to a forearm vein and connected by a shunt which could be removed to allow arterial blood to flow through the artificial kidney and be returned to the venous side. Between dialyses the shunt blood flow kept the connections patent. This technique, however, suffered from the fact that the shunt was externalized, necessitating an incision which frequently became infected. Introduction of a totally subcutaneous arteriovenous fistula by Brescia was a vast improvement (6). Such a fistula on a forearm artery to a vein produced a large venous flow which distended the vein, making it accessible to percutaneous puncture and eliminating the need for any external shunt. More recently arteriovenous fistulae have utilized vein autografts, bovine grafts, and woven Teflon® which becomes endothelialized.

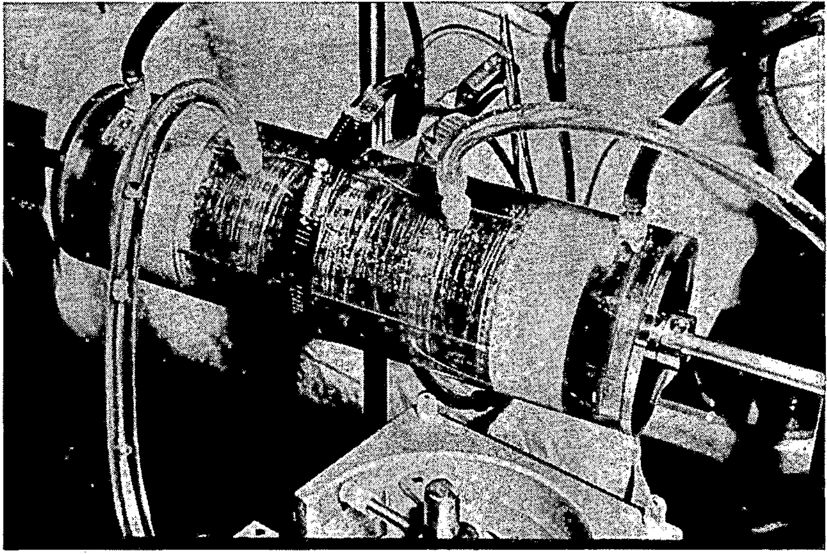


Figure 1 The hollow fiber artificial kidney.

Chronic dialysis can be performed in a hospital center for both inpatients and outpatients in an environment that is usually reserved for the sicker patients or those with more complications. Numerous satellite units for ambulatory patients have been set up where dialysis is largely accomplished by paramedical personnal under the supervision of a physician. About 25% of the 33,000 patients now on dialysis are treated in the home, with the help of either a family member or a nurse. In some instances the patient himself manages the whole operation. A few centers in which patients treat themselves in groups with a minimum of paramedical personnel have been set up and are functioning effectively. The economic advantages of such an operation are obvious.

Complications of Chronic Dialysis

Hemodialysis is an imperfect technique for treating end-stage renal failure. Apart from its inability to correct many of the metabolic deficiencies of the diseased kidney, a number of specific problems confront the patient on long-term dialysis. The death rate in the dialysis population is approximately 5–10% per year. Of these deaths the majority are due to infection or atherosclerotic disease, with stroke and coronary disease playing a major role (7). Hypertension certainly contributes to this atherosclerosis. Since many patients on chronic hemodialysis have hypertriglyceridemia, the latter was thought to contribute also, but recent evidence suggests that hypertriglyceridemia in the absence of elevation of serum cholesterol may only marginally affect progression of cardiovascular disease. Nevertheless, our inability to lower the triglyceride level in such patients is disappointing, particularly since the elevation may persist well into the posttransplant period. It is disturbing also that cardiovascular disease acquired during the dialysis period may be the cause of death in posttransplant patients whose kidney is functioning normally. Indeed the first successful kidney transplant patient, an identical twin, died about 7 years after surgery of a myocardial infarction at age 29.

Peripheral neuropathy characterized early by decrease in sensory nerve conduction and later by motor defects including foot drop and even paraplegia may develop in a patient with ESRD. If the onset is sudden and severe, adequate dialysis frequently causes marked improvement. In spite of months of adequate dialysis there is some question whether all the manifestations of neuropathy may be cured. A successfully functioning transplant, on the other hand, unquestionably "cures" neuropathy, but this may take as long as 6–8 months after normal renal function has been restored.

BONE DISEASE An early and almost invariable complication of renal failure may be manifested as osteomalacia, osteitis fibrosis cystica, osteosclerosis, or a combination of all three. Early in the course of chronic renal failure, the serum phosphate rises. This in turn depresses the serum ionized calcium, which stimulates the parathyroid and increases secretion of parathormone. Tubular resorption of phosphate decreases, the serum phosphate drops, ionized calcium rises, and as the level of both these ions returns towards the normal, they are maintained by continuing increased parathormone secretion, i.e. secondary hyperparathyroidism. In addition, since one of the most active metabolites of Vitamin D is formed in the kidneys, the characteristic Vitamin D resistance of chronic uremic patients contributes further to the bone disease. These events have pertinence for the dialysis patient who has received a successful transplant. Marked hypertrophy of the parathyroid glands may continue well into the posttransplant period with bone disease actually progressing (8). Steroid therapy in all probability contributes as well, although analysis of the various factors involved in our own series fails to implicate any single factor. Aseptic necrosis of the hip is common in the transplant clinic even among those patients with normal renal function (Figure 2). Of the patients in our series, 15% have aseptic necrosis of the hip that has required surgery. In one series reported from Holland, the incidence was as high as 40%. Although hyperfunction of the parathyroid under these circumstances has been called autonomous or tertiary, it is probable that there is some regulation of the markedly increased mass of parathyroid tissue by changes in serum ionized calcium. However, the amount of parathormone that continues to be released, even with normal renal function, results in continuing hyperparathyroidism and bone disease. Parathyroidectomy pre-, and even post-transplant is occasionally necessary. In recent years the active metabolite of Vitamin D, 1,25-dihydrocholecalciferol (1,25-DHCC), has become available. In clinical trials of patients on chronic dialysis, it has been shown to correct many of the defects in Vitamin D metabolism (9). Synthetic 1,25-DHCC will soon become available for general clinical use and should decrease the incidence of bone disease in ESRD patients. 1-α-hydroxycholecalciferol, a synthetic analog of Vitamin D, can bypass the block in renal hydroxylation present in chronic renal failure. It becomes effective when hydroxylated at the 25 position in the liver. Therapeutic trials with this agent, too, show improvement in the clinical manifestations of renal osteodystrophy (9). The use of aluminum hydroxide gels that bind phosphate in the gut may also delay the onset of secondary hyperparathyroidism, although it has recently been shown that significant amounts of aluminum can be absorbed from these substances and may contribute to neuropathy and central nervous system syndrome (10).

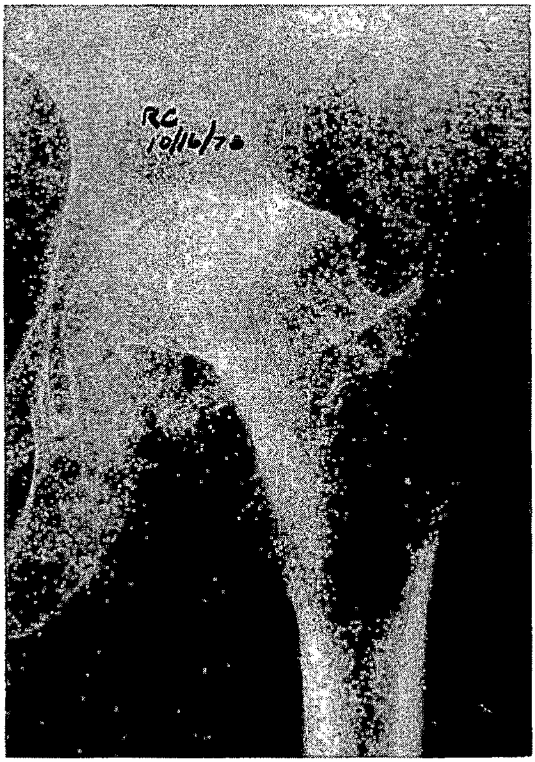


Figure 2 This X ray shows necrosis of the hip.

ANEMIA Anemia is a universal problem in ESRD patients on dialysis. Although folate may be lost in the dialysate bath, and iron absorbtion from the gut is poor in such patients, these two difficulties can be corrected by adequate replacement therapy. Unlike 1,25-DHCC, we have had until recently no active principle of erythropoieten that can be administered. Lack of erythropoieten production by the failing kidney remains the major cause of anemia. Interestingly, even remnants of kidney that cannot support life appear to maintain patients with higher hematocrits than occur in anephric patients. The hematocrit may stabilize at 20–25% in patients who are not transfused and then show a gradual rise. One suspects that such adaptation might result from an extrarenal source of erythropoieten, but this has not been demonstrated. In the consideration of dialysis versus transplantation in the treatment of anemia in ESRD, it is obvious that successful transplantation that promptly corrects anemia has a major advantage.

A hazard peculiar to dialysis is the contamination by various trace metals of water used to mix the dialysate. While systems for deionization and reverse osmosis have been devised to clear the water of contaminants, a number of reports have implicated base metals in the genesis of dialysis-induced disease. One of the earliest of these was copper intoxication presumed to be due to contact of the water with the heating coil (11).

HYPERTENSION Control of hypertension in dialysis patients is a major problem for the physician. Unquestionably the ingestion of sodium and water contributes to this in many patients and can be corrected by ultrafiltration. However, in some individuals, high levels of peripheral renin have been found and these patients do not respond as well to ultrafiltration. Indeed, their reaction to a weight loss of a kilogram or so by ultrafiltration may result in little change in supine blood pressure but in marked hypotension on assuming the upright position. Several years ago we believed that bilateral nephrectomy was the answer to patients with the hyperreninemia syndrome (12). However, careful management with some of the newer antihypertensive agents, particularly propanolol and minoxidil, have facilitated the medical management of such patients to the point where bilateral nephrectomy is rarely done now purely for hypertension. Autonephrectomy, effected by the injection of autologous fibrin clot into the renal artery via catheter introduced transfemorally, is a less invasive but probably less reliable technique for halting renin production. It is an interesting historical vignette that a form of treatment utilized thousands of years ago, proposed some ten years ago (13) for the treatment of ESRD, and revived only recently (14), may be of some benefit, particularly in facilitating losses of water and electrolytes between dialyses. The induction of sweating by the use of either a sauna bath or even a hot water bath may cause significant losses of sodium, water, and even urea. Following transplantation of a kidney that functions normally and in the absence of the patient's own kidneys, hypertension is rarely a therapeutic problem.

A third factor certainly compounds the vagaries of treatment of hypertension in the ESRD patient on dialysis. The vasomotor response of many such patients is abnormal, e.g. slowing of the heart in response to angiotensin infusions does not occur as it does in the normal, nor is this defect improved following dialysis (15) (Figure 3). The pupillary response to tyramine is abnormal and the response of blood pressure and heart rate with standard exercise is distinctly abnormal, even in "healthy" patients on chronic dialysis (16).

Psychosis and depression may occur in some patients who find intolerable the fact that their life depends upon being attached to a machine. This is not unique to dialysis patients, however, since many transplant patients live in constant anxiety about the possible failure of their allograft.

THE PROBLEM OF UREMIA What is uremia? In spite of what is generally considered adequate dialysis and rather moderate disturbances in the composition of body fluids, many patients fail to thrive. This is particularly true of the older age group. A detailed discussion of the long and unrewarding search for a uremic toxin is not within the scope of this paper. However, the most recent controversy should be mentioned. It was noted some time ago (17) that patients with peritoneal dialysis apparently felt better than patients adequately dialyzed with artificial kidney. A suggestion was therefore made that since the peritoneal membrane was relatively more permeable to larger molecules, that the toxin or toxins of uremia might reside in the so-called middle molecules—those of 1000–4000 daltons. A number of techniques utilizing different membranes,

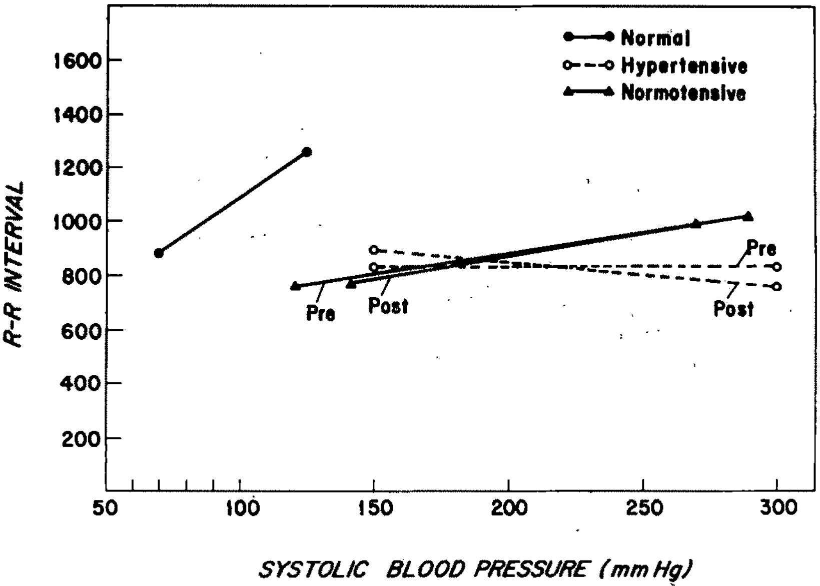


Figure 3 The effect of angiotensin infusions on pulse rate in normal and in uremic patients, both hypertensive and normal tensive, in the latter measured pre- and post-dialysis. It is seen that there is no slowing of the pulse rate with rise in blood pressure, as occurs in the normal, nor is this abnormal response affected by hemodialysis (15).

different areas, and different flows have been devised for more effective removal of these middle molecules. Fractionation of uremic sera using a DEAE-Sephadex microcolumn has elicited a peak containing middle molecules that appear to be correlated with the severity of the uremic syndrome (18). Nevertheless it can be said in summary that a number of careful trials using adequate controls have failed to prove that the removal of more middle molecules improves the treatment of the uremic syndrome (19). In part this may be because the objective methods of measuring such improvement are woefully inadequate.

SOCIOECONOMIC IMPACT OF TRANSPLANT AND DIALYSIS

Our ability to successfully transplant and dialyze a large number of patients since the early 1960s and the implementation of Congressional Bill HR-1 in October 1973 have had a profound impact on at least one segment of the patient population and the economy. First, HR1 provides for the funding of all patients on chronic dialysis and transplantation who are eligible for Social Security through the Medicare program. It is estimated that in 1977 there are 33,000 patients on dialysis in the United States and the combined cost of the transplantation and dialysis programs to the federal government will be 757 million dollars; by 1980 the projection is 45,000 patients on dialysis and a combined cost of 1.3 billion dollars. Careful evaluation of the program was undertaken by a Senate Committee in late 1976 and further implemented by hearings on a House Bill (HR 3112) in early 1977. Although the evaluation was in general well conceived and brought into clear focus the mechanisms for reviewing quality care and decreasing costs, it reflected a controversy that has existed for some time in the world of hemodialysis. A provision of the bill states that (20)

> the Secretary shall establish the appropriate proportion of individuals that should be referred to self-dialysis training and be engaged with self-care dialysis. With respect to all networks, such proportion shall be equal to 40% by October 1, 1978, 50% by October 1, 1980 and such additional percentage as the Secretary in consultation with the medical review board shall determine.

Home dialysis has obvious economic advantages (7,000–8,000 dollars per year) when compared with center dialysis (15,000 dollars) and limited service or self-dialysis units (12,000 dollars). However, the percentage of patients deemed suitable for home dialysis is hotly debated. Most nephrologists engaged in dialysis believe the figure to be closer to 20% than to 50%. In the Seattle area 85–90% of the patients are dialyzed at home. Probably 25% of our patients at the Peter Bent Brigham Hospital (PBBH) are on home dialysis. E. Friedman, the Director of Dialysis at Downstate Medical Center in Brooklyn, believes 70% of ESRD patients can be successfully trained for at-home dialysis. However, of 250 Downstate dialysis patients, only 60% actually dialyze at home. J. Soberman, Director of the nearby Bronx Dialysis Center, which treats a similar population, believes Friedman's estimate to be much too high. J. Soberman says that no more than 10% could be considered for at-home self-care (21). Obviously home dialysis is more difficult in the ghetto than in suburbia. Center dialysis has the very real advantage of affording an extra degree of security and considerable group psychotherapy. In my opinion limited service or self-dialysis units seem a reasonable compromise.

The second reason for the somewhat unexpected and staggering cost of dialysis in 1977 is the broadening of our indications for dialysis treatment of ESRD patients. The average age of the patient has increased considerably. Diabetics, even blind diabetics, are being treated, as are stable cancer patients such as those with very slowly progressive myeloma or stage I and II Hodgkin's disease. Does this broadening of our criteria reflect improvement in technology or possibly the fact that payment for such individuals is assured? The decrease in the number of patients on home dialysis suggests that the better Medicare coverage of center dialysis under HR 1 might account for this. However, an analysis of Figures 4 and 5 shows that the decrease of home dialysis patients began sometime before the implementation of HR 1 and that in all probability increasing patient age may make them less suitable candidates.

TRANSPLANTATION

In the last decade our knowledge of the immunology of rejection of transplanted tissues has increased tremendously. With the advent of immunosuppressive therapy, some 27,000 patients have received kidney transplants. One of the factors that contributes to the success of any tissue or organ allograft between

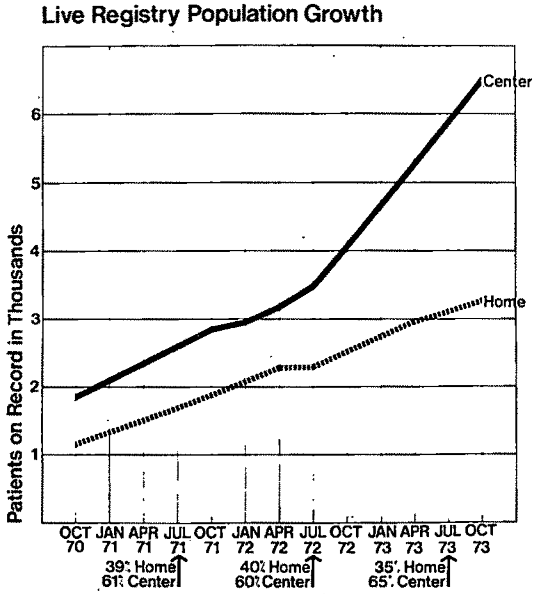


Figure 4 The proportion of patients on home dialysis to those on center dialysis began to decrease sharply before the implementation of HR1 in October 1973.

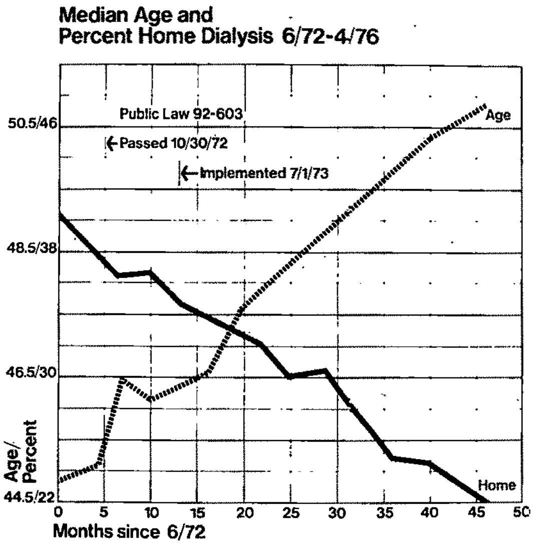


Figure 5 The median age of patients and the percent of those on home dialysis are plotted against each other. As the patient age increases, the percentage of those on home dialysis decreases.

two individuals is their tissue compatability. Even as a blood transfusion can be successfully given if the blood is of the same type, tissues can be transplanted between humans if the tissue types are identical. Typing of human tissues has become an important endeavor not only for transplantation but because of the relationship of various tissue types to certain diseases. Some 38 tissue types are now recognized. Each individual has 4 major tissue types, 2 on the chromosome from the father and 2 from the mother (22). In addition, there are various other loci on the chromosome near the tissue type (HLA) locus that may determine compatibility, immunoreactivity to various forms of antigen stimulus, and the propensity for the development of particular disease syndromes (23). There is a close relationship between tissue type and ankylosing spondylitis. Even more exciting is the relationship between juvenile diabetes and tissue type and ancillary evidence, which suggests that juvenile diabetes may be an inherited susceptibility to viral attack on the islets of Langerhans (24).

In spite of the advances in our knowledge of transplantation immunology, the clinical results over the past five years have been disappointing. Table 1 shows that, at least in the transplantation of cadaver kidneys, our results have improved very little. Table 2, however, shows that, while the survival of the kidney has not improved, the patient survival has. These figures are universally acknowledged to mean that our present immunosuppressive therapy is inadequate and that attempts to press immunosuppressive therapy in the face of rejection will make the patient more susceptible to infection and the other dire consequences of this regimen. The philosophy of saving the patient at the expense of

Table 1 Functional survival first transplant

Donor source	Year of transplant	Sample size	1 yr. (%)	2 yr. (%)
Cadaver	1968	649	47.6	40.4
	1969	847	54.2	47.2
	1970	1147	55.3	47.2
	1971	1559	53.1	45.7
	1972	1707	50.6	42.6
	1975		50	42

the kidney if necessary has undoubtedly contributed to this decrease in mortality. Furthermore, rejection of a first kidney does not appear to jeopardize the success of a second transplant provided that specific sensitization has not occurred. The improvement in mortality in patients receiving cadaver transplants resolves a long-standing argument between the dialyzer physician and the transplanter. As noted above, mortality on dialysis is some 5–10%. This is slightly more than the risk taken by a patient who has been well dialyzed and then receives a cadaver kidney transplant. Our experience at PBBH is illustrative; we have lost no patients of the last 70 receiving cadaver transplants.

The extensive advances in our knowledge of the immunology of transplantation again is beyond the scope of this paper. However, a few significant points must be mentioned. After a long trial period of antilymphocyte globulin, a therapy whose efficacy was debated, the role of the probably more specific antithymocyte globulin is now established. The use of antilymphocyte globulin (ALG) resulted from the early work of Woodruff (25), who reasoned that antisera against human lymphocytes might destroy recipient lymphocytes and thus lessen the immunologic assault on transplanted tissue. Various types of ALG were raised by immunizing horses, rabbits, and goats with human lymphocytes. The results were equivocal. However, more recently the use of human thymocytes removed from children at the time of cardiac surgery and utilized to induce antibody in rabbits has apparently made a considerable difference.

Corticosteroid therapy for rejection previously given by mouth has now been supplemented by quite a different method. Corticosteroid as an immunosuppressive agent acts within a matter of days. For this reason, it is now given as

Table 2 Patient survival of first transplant

Donor source	Year of transplant	Sample size	1 yr. (%)
Cadaver	1968	649	58.6
	1969	847	65.5
	1970	1147	69.2
	1971	1559	69.2
	1972	1707	71.8
	1975		85–90

"pulse" therapy. When rejection crisis is suspected, massive doses of intravenous methylprednisolone (as much as 1.5–2 g) are given for 3–4 days; the dose is then decreased rapidly and given orally. This avoids many of the dire side effects of steroid therapy and is apparently as effective as more prolonged doses of oral steroids. In the transplant program at PBBH we have markedly diminished the problem of infection by the use of gentamicin and oxacillin given in large doses intraoperatively and for one day thereafter. While this apparently violates the caveat of prophylactic antibiotic therapy, our own results and the experience of others with short-term prophylaxis apparently justifies this approach to treatment.

TRANSFUSION

The linkage of dialysis and transplantation is nowhere better exemplified than in the role played by transfusion of the patient on chronic dialysis. Dialysis patients may require multiple transfusions. Transfusions of whole blood and even frozen blood contain elements of white cells that carry the human HLA antigen—as such they can sensitize or immunize the recipient to this antigen. This is equally true of multiparous women who may be sensitized by the antigens of the fetus bearing their spouse's genes. It is unquestionably true that a recipient with antibodies specific for the antigens of the donor will have an accelerated rejection of the transplanted kidney. However, to our great surprise, dialysis patients who have received no transfusions do less well on receiving a kidney than do patients who have had multiple transfusions but whose antibodies are not specific for the donor. This fact raises the interesting and very important question of whether the latter sera contain blocking antibodies as well, and remains an important consideration for the future.

HEPATITIS

The problem of hepatitis remains an enigma in the dialysis-transplant endeavor. Type A hepatitis virus certainly has no role in the transmission by blood transfusion. However, type B hepatitis virus unquestionably transmits disease by transfusion. Unfortunately, there are almost certainly several other viruses not classified by A and B that constitute 20% of transfusion-transmitted hepatitis. Catastrophic episodes of hepatitis have occurred in dialysis units, notably that at the University of Edinburgh, but these are unusual. A recent report from Belgium suggests that patients positive for hepatitis B antigen (HB_sAg) have a greater incidence of chronic liver disease than those who are not positive (26). This has not been the experience in the United States (27). However, most nephrologists would agree that a patient positive for one of the hepatitis B antigens need not necessarily be refused transplantation or chronic dialysis. Interestingly, allograft recipients with HB_sAg appear to tolerate their grafts better, while recipients with anti-HB_sAg reject the graft; this suggests perhaps a difference in immunologic activity of the two groups (28).

COMPLICATIONS OF TRANSPLANTATION

Even as the indications for chronic hemodialysis have expanded, those for transplantation have enlarged. Many diabetics have been transplanted; the prognosis is poorer than with non-diabetics but some have survived for many useful years. A study comparing dialysis and transplantation of diabetic patients concluded that "rehabilitation and long term survival favored transplantation as the treatment of choice for diabetic patients with ESRD" (29). Transplants of pancreatic islet cells have been attempted and there is some evidence that they will survive in the liver and reduce the insulin requirement for months or more. This remains to be evaluated. The kidney, of course, is the source of a number of enzymes and has been used as an enzyme replacement in several diseases characterized by congenital enzyme deficiency such as Fabrys and oxalosis. Unfortunately, in neither of these syndromes has this procedure been successful.

Recurrent Nephritis

Recurrence of the same type of nephritis that destroyed the kidney in the original recipient takes place in approximately 15% of kidney transplants. In the specific histologic type "dense deposit disease" recurrence has been reported as high as 40%. However, recurrence of the disease may be manifested merely as proteinuria and the prognosis may not be as bad as was anticipated in earlier reports.

Malignancies

It has been known for some time that there is an increased incidence of malignancies in patients who had received immunosuppression and renal allograft. The best data available are those collected by Penn (30), which showed an incidence of 5.6% of cancer in such patients, 100 times greater than that observed in the general population within the same age range. The types of cancer were somewhat unusual; most prominent were cancers of the skin and lips, lymphomas, and carcinoma of the cervix. Particularly noteworthy was the frequency with which the central nervous system was involved in lymphomas. One would suspect, if the frequently quoted postulate that the immunosurveillance mechanism for transformed cells were blunted or destroyed by immunosuppression, that cancer of the breast in women and cancer of the prostate in men would be most common, but this is not the case. Lest dialysis in ESRD should be unencumbered by the stigma of the development of neoplasia, in the recent report by Slifkin (31) dialysis patients had an incidence rate of cancer of approximately 5 times that of a control population when matched for age and all sites. Although the analysis of patients surveyed represented a smaller total number than those of the transplants surveyed, the cancers included a much wider distribution.

Gastrointestinal bleeding

A continuing problem in transplantation is gastrointestinal bleeding. Patients with preoperative duodenal ulcers frequently bleed after transplantation as do patients with polycystic disease who often have diverticulae. A number of surgeons in the transplantation area will do partial gastrectomies and vagotomies prior to transplantation in patients who have had gastric or duodenal ulcers. Obviously the immediate approach for the future is that of more specific immunosuppressive therapy, i.e. suppressive therapy only against the graft, not against invading organisms from which the patient may die (32). From animal experiments there is much evidence that this may become possible.

Several examples of this approach may be cited as guideposts to future treatment of recipients of human renal allografts. The work of Brent and his colleagues (33) has demonstrated that it is possible to produce true "specific immunologic unresponsiveness," meaning that the recipient of a skin graft properly pretreated will retain the skin graft from the specific donor but reject those from all other donors. Obviously this leaves the immune system available for response to invasion by microorganisms. Timing and dosage of the agents utilized in producing such unresponsiveness are critical, however. Whether such complicated maneuvers could be applied to humans is problematical but certainly possible. A second approach, which has already been demonstrated to be effective in dogs, is the use of the $F(AB)^2$ fragments of specific antibody to the graft raised in a second animal; since the $F(AB)^2$ fragment does not bind complement, it is not destructive. Nevertheless, it apparently does block the binding site for whole antibody and thus spares the graft. Since it is possible to utilize human antisera and defractionate it in this fashion, this approach may well be feasible for humans in the near future (34). Finally, Strom and his coworkers (35) in our laboratory have demonstrated that the "killing effect" of recipient lymphocytes already sensitized to donor antigens may be markedly decreased by the use of two agents in common use in clinical therapy today: isoproterenol and aminophyllin. Although this is an in vitro phenomenon, the administration of these agents to rat recipients of renal allografts has been shown to prolong survival of the grafted kidney.

Ultimately, however, the prevention or arrest of glomerulonephritis, which constitutes two thirds of the cause of ESRD, will be the answer. This is a first-order technology as opposed to dialysis and transplantation, which are second-order technologies. The analogy is that of the iron lung (dialysis and transplantation) vs polio vaccine (prevention of ESRD).

SUMMARY

Both dialysis and kidney transplantation are effective techniques for prolonging life in ESRD. Because dialysis therapy does not effect replacement of the metabolic functions of the kidney, it is less than perfect. Successful transplantation that replaces all of the aspects of renal function is the treatment of choice. Successful transplantation is highly dependent upon the availability of a suitable donor and the appropriate tissue match, which remains a problem.

Present immunosuppressive therapy required to prevent the immunoresponse causing rejection of the renal allograft is a tool too dull for the job. Since all immunoresponse is suppressed, infections are common, as are the multiple complications of cortical steroid therapy. For the dialysis patient, development of more compact effective dialysis apparatus and particularly the availability of replacement therapy hold promise. New approaches to diminishing the immune response to the graft without impairing that to microorganisms may well effect improvement in graft survival as will increasing knowledge of factors other than HLA antigens in the immunologic reaction. Development of an effective method for arresting the progress of glomerulonephritis before it reaches end-stage renal failure would obviate the necessity for dialysis or transplant therapy in approximately two thirds of ESRD patients.

Literature Cited

1. Hume, D. M., Merrill, J. P., Miller, B. F., Thorn, G. W. 1955. Experience with renal homotransplantation in the human. Report of 9 cases. *J. Clin. Invest.* 34:327
2. Tenckhoff, H. 1974. Peritoneal dialysis today—a new look. *Nephron* 12:420
3. Friedman, E. A., Blagg, C. R., Sullivan, J. P., Briefel, R. S., Galonsky, P. S., Delano, B. G., Hutchinson, J. T. 1977. *Trans. Am. Soc. Artif. Intern. Organs* 6:24 (Abstr.)
4. Kolff, M. D. 1976. Open Letter, *Dialysis and Transplantation*, April/May, p. 80
5. Quinton, W., Dollard, D., Scribner, B. H. 1962. Cannulation of blood vessels for prolonged hemodialysis. *Trans. Am. Soc. Artif. Intern. Organs* 8:315
6. Brescia, M. J., Cimino, J. E., Appel, K., Hurwich, B. J. 1966. Chronic hemodialysis using venipuncture and surgically created arteriovenous fistula. *N. Engl. J. Med.* 275:1089–92
7. Lowrie, E. G., Lazarus, J. M., Hampers, C. L., Merrill, J. P. 1974. Cardiovascular disease in dialysis patients. *N. Engl. J. Med.* 290:737–38
8. Gottlieb, M. N., Stephens, M. K., Lowrie, E. G., Lazarus, J. M., Griffiths, H. J., Kenzora, J. E., Strom, T. B., Philips, E., Merrill, J. P. 1977. In *Calcium Metabolism in Renal Failure and Nephrolithiasis*, ed. D. S. David, pp. 279–91. New York: Wiley
9. David, D. 1977. Mineral and bone hemeostasis in renal failure. In *Calcium Metabolism in Renal Failure and Nephrolithiasis*, ed. D. S. David, Chap. 1. New York: Wiley
10. Alfrey, A. C., Legendre, G. R., Kaehney, W. D. 1976. *N. Engl. J. Med.* 294:184
11. Klein, W. J. Jr., Metz, E. N., Price, A. R. 1972. Acute copper intoxication; a hazard of hemodialysis. *Arch. Intern. Med.* 29:578
12. Lazarus, J. M., Hampers, C. L., Merrill, J. P. 1974. Hypertension in chronic renal failure. Treatment with hemodialysis and nephrectomy. *Arch. Intern. Med.* 133:1059–65
13. Snyder, D., Merrill, J. P. 1966. Sauna baths in the treatment of chronic renal failure. *Trans. Am. Soc. Artif. Intern. Organs* 12:188
14. Sacher, J. W., Schrier, R. W. Sweating treatment for chronic renal failure. To be published
15. Lazarus, J. M., Hampers, C. L., Lowrie, E. G., Merrill, J. P. 1973. Baroreceptor activity in normotensive and hypertensive uremic patients. *Circulation* 47:1015–20
16. Milne, F. J., Guz, A., Murphy, K., Purcell, H. J., Coppin, K., de Wardener, H. E. The exercise responses of healthy hemodialysis patients. PhD thesis, Univ. Cape Town (Charing Cross Hospital, London
17. Scribner, B. H. 1965. Discussion. *Trans. Am. Soc. Artif. Intern. Organs* 11:29
18. Bergstrom, J., Furst, P., Gordon, A. 1976. A study of uremic toxicology. *Proc. 8th Annu. Contractors Conf.* Natl. Inst. Arthritis, Metab., Digestive Dis., DHEW Publ. No. (NIU) 76-248
19. Gotch, F. A. 1976. Urea guided dialysis therapy. Current clinical results. *Dialysis and Transplantation*, Oct./Nov., p. 15

20. A bill to amend titles II and XVIII of the Social Security Act to make improvements in the end-stage renal disease program presently authorized under section 226 of that Act. 95th Congress HR 3112. Mr. Rostenkowski and Mr. Vanik. Feb 3, 1977
21. Do-it-yourself dialysis need not be done at home. *Med. World News,* Apr. 18, 1977. p. 34
22. Carpenter, C. B., Merrill, J. P. 1977. Transplantation. In *Principles of Internal Medicine*, pp. 413–26. NY: McGraw Hill. 8th ed.
23. VanRood, J. J. 1975. Disease predisposition, immune responsiveness and fine structure of the HL-A supergene. *Transplant. Rev.* 22:75
24. Svejgaard, A. 1975. HL-A and disease associations—a survey. *Transplant. Rev.* 22:3
25. Woodruff, M. F. A., Anderson, M. F. 1963. Effects of lymphocyte depletion by thoracic duct fistula and administration of antilymphocyte serum on the survival of skin homografts in rats. *Nature* 200:702
26. Pirson, Y., Alexandre, G. P. J., de Strihou, C. VanY. 1977. Long-term effect of HB_s antigenemia on patient survival after renal transplantation. *N. Engl. J. Med.* 296:194–96
27. Strom, T., Merrill, J. P. 1977. Hepatitis B. transfusions and renal transplantation. *N. Engl. J. Med.* 296:225–26
28. Rashid, A., Sengar, D., Couture, R., Jindal, S., Posen, G. 1977. Relationship of hepatic B infection to graft survival in cataveric renal transplantation. *Trans. Am. Soc. Artif. Organs* 6:71 (Abstr.)
29. Rao, K. V., Sutherland, D., Kjellstrand, C. M., Shapiro, F. L. 1977. Comparative results between dialysis (D) and Transplantation (T) in diabetic patients. *Trans. Am. Soc. Artif. Intern. Organs* 6:70 (Abstr.)
30. Penn, I. 1976. Incidence of malignancies in transplant recipients. *Transplant. Proc.* (2), 7:323
31. Slifkin, R. F., Neff, M., Baez, A., Gupta, S., Matoo, N. 1977. Malignancy in end stage renal disease. *ASAIO* 6:81 (Abstr.)
32. Merrill, J. P. 1977. An evaluation of immunosuppressive treatment for human kidney allografts. *Seminaires d'uro-nephrologie,* p. 80. Paris, New York: Masson
33. Brent, L., Pinto, M. 1974. Induction of specific unresponsiveness by donor antigen and non specific immunosuppression. In *Immunological Aspects of Surgery Transplantation*
34. Wren, S. F. G., Martins, A. C. P., Von Haefen, U., Busch, G. J., Katske, G. E., Koppenheffer, T. L., Shaipanich, T., Wilson, R. E. 1974. Passive enhancement of canine renal allografts with polyspecific F(AB')2 fragments. *Surgery* 76:112–20
35. Strom, T. B., Carpenter, C. B., Garovoy, M. R., Austen, K. F., Merrill, J. P., Kaliner, M. 1973. The modulating influence of cyclic nucleotides upon lymphocytemediated cytotoxicity. *J. Exp. Med.* 138:381–93

Ann. Rev. Med. 1978. 29:359–66
Copyright © 1978 by Annual Reviews Inc. All rights reserved

NON-A, NON-B HEPATITIS ◆7285

Stephen M. Feinstone, M.D. and Robert H. Purcell, M.D.
Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Disease, National Institutes of Health, USPHS, DHEW, Bethesda, Maryland 20014

INTRODUCTION

As a result of the development of sensitive serologic tests for hepatitis B virus (HBV) and hepatitis A virus (HAV)[1] infection, it became clear that there were certain patients with hepatitis who had no evidence for infection by either HAV or HBV (1). These serologic findings were in contrast to early epidemiologic and volunteer studies, which consistently provided evidence for only two types of viral hepatitis (2). However, occasional epidemiologic studies reported patients with more than two distinct episodes of hepatitis (3–5), and in a recent well-documented report, Mosely (6) found patients with three and four distinct bouts of hepatitis. Furthermore, analyses of the incubation periods of hepatitis cases following blood transfusion revealed that they do not conform to a bimodal curve, as would be expected if only types A and B hepatitis were occurring. These observations, and others to be discussed later, led to the concept that another as yet undetected agent or agents can cause human viral hepatitis.

KNOWN VIRAL AGENTS OF NON-A, NON-B HEPATITIS

Epstein-Barr virus (EBV) infections are commonly associated with hepatitis as part of the infectious mononucleosis syndrome. However, this type of hepatitis can usually be diagnosed clinically as well as with routine laboratory procedures. In addition, there frequently are characteristic pathologic changes

---

[1] HBV has several morphologic forms and antigenic determinants associated with it. The Dane particle is thought to be the actual HBV. It is a complex virus-like structure about 45 nm in diameter with an outer coat, which contains the antigenic determinants of the hepatitis B surface antigen (HBsAg), and an inner core of approximately 27 nm, which contains the hepatitis B core antigen (HBcAg). HBsAg also exists as spheres or filamentous structures with diameters of about 22 nm. The HBsAg is antigenically identical to the Australia antigen first detected by agar gel diffusion by Blumberg, et al (40). Antibody to these antigens are termed anti-HBs and anti-HBc. The hepatitis A antigen (HA Ag) is a 27-nm virus-like particle, first detected in stools from infected individuals. Its antibody is termed anti-HA (41).

0066-4219/78/0401-0359$01.00

found on liver biopsy, which are helpful in making the diagnosis (7). In several studies of posttransfusion hepatitis, EBV has rarely been implicated in cases of non-B hepatitis, as judged by seroconversions (1, 8–12).

The role of cytomegalovirus (CMV) in non-A, non-B hepatitis is more difficult to evaluate because of the presence of multiple, partially related serotypes of the virus and a demonstration of variation in anti-CMV antibody levels in normal individuals followed longitudinally (13, 14). Convincing evidence for an etiologic role of CMV in some cases of hepatitis associated with transfusion or immunosuppression has been presented (15, 16), but most cases of non-B hepatitis lack such evidence. In addition, serologic evidence of recent CMV infection can be found with equal frequency in transfused patients who develop no hepatitis, type B hepatitis, or non-B hepatitis (8, 17–19). Thus, the importance of CMV in the etiology of non-B transfusion-associated hepatitis is difficult to evaluate but is probably not great.

Although other viruses, notably enteroviruses, have been reported to be associated with individual cases of hepatitis, these reports must be interpreted with caution, as many of these viruses are very common, and their detection during a bout of hepatitis could easily be coincidental. Well-documented cases of hepatic infection with enteroviruses exist; however, the hepatic involvement occurs as part of a more generalized enterovirus infection (20).

EPIDEMIOLOGY

Until tests became available for type B hepatitis, cases were diagnosed as type A or type B according to epidemiologic and clinical criteria. The early, insensitive tests for HB did little to change this, because cases that did not fit the accepted patterns were felt to be misdiagnosed serologically. With highly sensitive tests for HBsAg and anti-HBs, long-held epidemiological concepts had to be changed (6). In an early study of transfusion-associated hepatitis (8) an etiologic agent could not be found for over half the cases. It was assumed that many of these cases were caused by HBV but could not be diagnosed because of the insensitivity of the tests. However, few additional HBV infections were identified when more sensitive serologic tests were applied to serial samples from these patients (R. H. Purcell, unpublished).

Non-B posttransfusion hepatitis has been distinguished from type A hepatitis on both epidemiologic and serologic grounds (17, 1). It has now been shown that HAV infections are rarely transmitted by blood transfusion. Hepatitis A accounts for most hepatitis occurring in those epidemic situations usually related to contaminated water or food and documented to occur within institutions, families, and among handlers of nonhuman primates (21, 22). HAV is endemic in many underdeveloped countries and infects most members of the population in childhood (23). Small epidemics following exposure to a common source of blood or blood products (such as a single lot of plasma or a contaminated syringe) are usually not related to hepatitis A. Sporadic (nonepidemic) hepatitis cases may be caused by HBV, HAV, or non-A, non-B hepatitis agents (24).

Non-A, non-B hepatitis appears to have a worldwide distributon. It has been detected essentially anywhere it has been carefully looked for including the United States (1), Japan (Y. Moritsugu, personal communication), Costa Rica (25), Australia (Gust et al, in preparation), Great Britain (26), and Germany (27).

As with type B hepatitis, transfusion-associated non-A, non-B hepatitis occurs more frequently when the blood is obtained from commercial rather than voluntary sources. If only voluntary blood is transfused, the rate of posttransfusion hepatitis not due to HBV can be reduced by five- to tenfold compared to the rate if commercial blood is used. However, if only voluntary blood is transfused and HBsAg-containing blood is screened out by the most sensitive techniques, hepatitis still occurs at a rate of about 3–6 cases/1000 units of blood transfused. In different studies 80–100% of this posttransfusion hepatitis (PTH) is non-B and presumably non-A (28, 29).

In summary, the epidemiology of non-A, non-B hepatitis resembles that of type B hepatitis more closely than that of type A disease. Non-A, non-B hepatitis commonly occurs following parenteral exposure to blood, is more common in lower socioeconomic groups such as commercial blood donors, and does not appear to be spread frequently from person to person by direct contact. Based on the observed rate of disease following transfusion, an infectious carrier state must exist in relatively asymptomatic people.

CLINICAL CHARACTERISTICS

The average non-A, non-B hepatitis case appears to differ clinically in several small but significant ways from type B illness, but as with hepatitis B, the clinical expression of non-A, non-B hepatitis is so variable that a diagnosis on clinical grounds alone cannot be made.

The incubation periods observed in non-A, non-B hepatitis have a wide range, which overlaps the incubation periods of both types A and B hepatitis. However, in the non-A, non-B cases we studied, the incubation periods averaged six or seven weeks shorter than type B cases (Table 1).

There seems to be general agreement that the acute phase of non-A, non-B hepatitis is on the average less severe than hepatitis B (Table 2). The variability of the disease is such that the evidence for hepatitis may be as little as mildly elevated enzymes, while other cases have typical icteric hepatitis, or fulminant

Table 1 Transfusion-associated hepatitis: incubation period

Type of hepatitis	Number	Mean incubation period (wk)[a]	Range (wk)
B[b]	18	14.4	4–25
Non-A, non-B	22	7.3	2–22

[a] From transfusion to first elevated SGPT value.
[b] HBsAg positive.

Table 2 Transfusion-associated hepatitis: severity

Type of hepatitis	Number	Number icteric	Mean maximum SGPT value	Range
B	19	12 (63%)	768	157–2050
Non-A, non-B	22	6 (27%)	395	120–1476

hepatitis. Furthermore, we and others have noted that the clinical expression of non-A, non-B hepatitis is often episodic. Patients may have two or more periods of illness manifested by enzyme elevations separated by periods of normal enzyme levels (Figure 1). Care must be taken that these repeating episodes not be interpreted as different bouts of hepatitis due to different infections.

Although non-A, non-B hepatitis is usually not as severe in the acute state as HB, it causes very similar chronic manifestations of infection. Indeed, chronic hepatitis may occur much more frequently in non-A, non-B disease (see Table 3). In some patients we have documented elevated enzymes lasting several years. In a recent study of 44 patients with non-A, non-B posttransfusion hepatitis, 10 developed chronic hepatitis. On liver biopsy one had chronic persistent hepatitis, eight had chronic active hepatitis, and one had cirrhosis (30). It is probable that a significant proportion of chronic hepatitis not related to HBV or other specific etiologies is initiated by infection with the agent(s) of non-A, non-B hepatitis. Because of this potential for producing chronic liver disease, control of non-A, non-B hepatitis is a major concern.

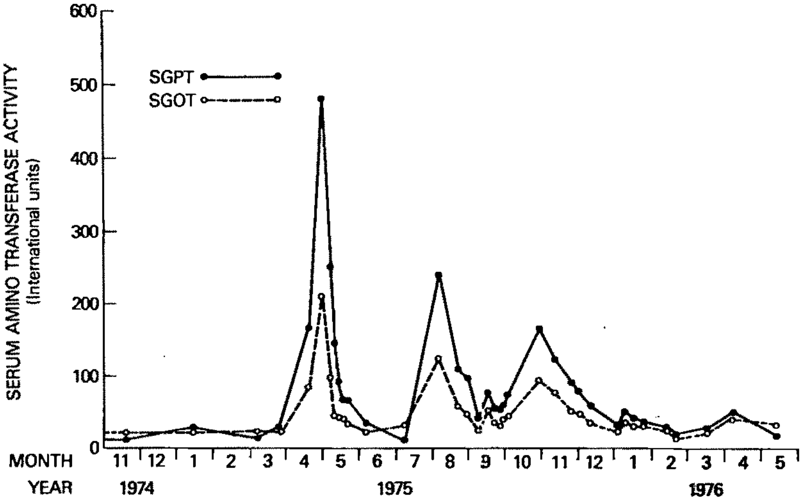


Figure 1 Serial serum enzyme levels in a nurse who developed non-A, non-B hepatitis following exposure to a single patient with a similar disease. There are three distinct periods of elevated enzymes during one year in this nurse, who had a 10-year record of normal enzyme levels. A biopsy obtained during the last period of enzyme elevation was read as resolving hepatitis.

THE CAUSE OF NON-A, NON-B HEPATITIS

Although intensive research is underway in several laboratories, at the time of this writing there has been little progress in identifying the specific agent or agents that cause non-A, non-B hepatitis. No culturable agent or related antigen has been detected. More surprising is the fact that no animal model for this disease has been found. Chimpanzees are susceptible to infection with both types A and B hepatitis viruses and respond with illnesses similar to hepatitis in man. However, attempts to infect chimpanzees or other nonhuman primates have to this date yielded mixed results. Since there are no antigenic markers for this agent(s), it is difficult to state whether the various species inoculated simply are not susceptible to infection by the agent of non-A, non-B hepatitis, have had prior exposure and are therefore immune, or if the various inocula tested contained infectious material.

One cause of transmissible hepatitis in marmoset monkeys, the GB agent, has been under investigation for several years and must be considered a candidate non-A, non-B hepatitis virus. The GB agent was first described by Deinhardt et al (31). They reported the serial transmission in marmosets of an agent (GB) derived from a surgeon who was thought to have hepatitis A. Parks & Melnick (32, 33) cast doubt on the relationship of GB to hepatitis A and suggested that it was in fact a latent marmoset agent. Deinhardt later showed that marmosets were susceptible to bona fide HAV that was not related to GB as revealed by cross-challenge experiments (34). It is possible that the original patient indeed had a form of non-A, non-B hepatitis that can be transmitted to marmosets. There are however, no serologic tests for the GB agent other than cross-challenge studies in marmosets. Therefore, it is still not possible to state the nature of the GB agent or what, if any, relationship it has to human hepatitis.

EVIDENCE FOR MORE THAN ONE NON-A, NON-B HEPATITIS AGENT

Mosley recently reported six patients who had three distinct bouts and five patients with four bouts of hepatitis (6). Serologic analysis of these multiple bouts revealed that as many as two or three separate bouts of hepatitis in one individual were not identifiable as type A or B and were therefore classified as non-A, non-B hepatitis (35). These bouts did not appear to be instances of the recurrent, chronic non-A, non-B hepatitis discussed above, because liver biopsy revealed each of them to be associated with histopathologic evidence of acute hepatitis only. Both EBV and CMV were excluded serologically as

Table 3 Transfusion-associated hepatitis: duration of illness

Type of hepatitis	Number	Number with SGPT elevation persisting >6 months
B	19	3 (6%)
Non-A, non-B	21	7 (33%)

causative agents in any of these cases. Without a direct serologic test for any of the infectious agents, it cannot be stated with certainty that these separate episodes of hepatitis represent separate infections. However, these findings plus the variability of the epidemiologic patterns, incubation periods, and clinical expression of illness discussed above all point to the possibility of two or more agents causing non-A, non-B hepatitis.

PREVENTION

At present the most effective way of preventing transfusion-associated non-A, non-B hepatitis is by the exclusive use of blood derived from voluntary sources. This measure has been shown to reduce markedly (but not eliminate) the incidence of type B and non-A, non-B hepatitis (36).

Several studies of the efficacy of immune serum globulin for the prevention of posttransfusion hepatitis have yielded conflicting results. Minimal, if any, prevention of non-A, non-B hepatitis by immune serum globulin was demonstrated in the Veterans Administration Cooperative Posttransfusion Hepatitis Study, which separated type B from non-B (presumably non-A) posttransfusion hepatitis (28). However, in another study conducted by the Army, gamma-globulin not only reduced the frequency and severity of non-A, non-B posttransfusion hepatitis, but it also reduced the frequency of chronic liver disease in the patients who did develop non-A, non-B hepatitis (30, 37).

Until there are direct means of detecting the non-A, non-B hepatitis agent(s) and its carriers, nonspecific tests might prove useful for preventing infection by blood transfusion. One possibility currently under investigation is to screen blood donors by testing serum enzyme levels. Although previous studies showed enzyme testing was not worthwhile in preventing PTH (38), now that the bulk of HBV carriers have been eliminated as blood donors, a correlation may be found between abnormal liver enzymes and the ability to transmit non-A, non-B hepatitis. A recent preliminary report indicated that carcinoembryonic antigen (CEA) was frequently found in blood transfused to patients who subsequently developed non-A, non-B hepatitis. In addition, CEA was often detected in the recipient's blood during these particular episodes of hepatitis. Some patients who subsequently developed chronic hepatitis also became chronic carriers of CEA (39). These types of indirect tests must be carefully evaluated in a large study of transfusions before the expense of performing them routinely can be justified.

CONCLUSION

There is evidence that at least one additional agent is etiologically responsible for human viral hepatitis and that this agent is antigenically unrelated to hepatitis A or B viruses. The epidemiology of such non-A, non-B hepatitis resembles that of type B hepatitis more than that of type A disease. Non-A, non-B hepatitis frequently follows blood transfusion and can progress to a chronic state, which makes its control highly important. Attempts to identify an antigen specific for non-A, non-B hepatitis or to transmit the disease to nonhuman primates have, to date, been unsuccessful. At present, the only known means to prevent this disease is by the exclusive use of volunteer sources of blood for transfusions. The efficacy of immune serum globulin in preventing non-A, non-B posttransfusion hepatitis is poorly defined. Other nonspecific methods for screening blood donors such as measuring enzyme levels are currently under investigation.

NOTE ADDED IN PROOF Studies recently completed (H. J. Alter et al; E. Tabor et al) report the transmission of non-A, non-B hepatitis to chimpanzees. Infection was documented by both biochemical and histological evidence of liver disease.

Literature Cited

1. Feinstone, S. M., Kapikian, A. Z., Purcell, R. H., Alter, H. J., Holland, P. V. 1975. Transfusion associated hepatitis not due to viral hepatitis type A or B. *N. Engl. J. Med.* 292:767–70
2. Krugman, S. 1976. Viral hepatitis: Overview and historical perspectives. *Yale J. Biol. Med.* 49:243–50
3. Havens, W. P. 1956. Viral hepatitis: Multiple attacks in a narcotic addict. *Ann. Intern. Med.* 44:199–203
4. Levine, R. A., Payne, M. A. 1960. Homologous serum hepatitis in youthful heroin users. *Ann. Intern. Med.* 53:164–78
5. Rosenstein, B. J. 1967. Viral hepatitis in narcotic users: An outbreak in Rhode Island. *J. Am. Med. Assoc.* 199:698–700
6. Mosley, J. W. 1975. The epidemiology of viral hepatitis: An overview. *Am. J. Med. Sci.* 270:253–70
7. Scheuer, P. J. 1973. *Liver Biopsy Interpretation,* p. 133. Baltimore: Williams & Wilkins, 171 pp.
8. Purcell, R. H., Walsh, J. H., Holland, P. V., Morrow, A. G., Wood, S., Chanock, R. M. 1971. Seroepidemiological studies of transfusion-associated hepatitis. *J. Infect. Dis.* 123:406–13
9. Gocke, D. J. 1972. A prospective study of posttransfusion hepatitis: The role of Australia antigen. *J. Am. Med. Assoc.* 219:1165–70
10. Alter, H. J., Holland, P. V., Purcell, R. H., Lander, J. J. Feinstone, S. M., Morrow, A. G., Schmidt, P. J. 1972. Posttransfusion hepatitis after exclusion of commercial and hepatitis B antigen-positive donors. *Ann. Intern. Med.* 77:691–99
11. Alter, H. J., Purcell, R. H., Holland, P. V., Feinstone, S. M., Morrow, A. G., Moritsugu, Y. 1975. Clinical and serological analysis of transfusion-associated hepatitis. *Lancet* 2:838–41
12. Knodell, R. G., Conrad, M. E., Dienstag, J. L., Bell, C. J. 1975. Etiologic spectrum of posttransfusion hepatitis. *Gastroenterology* 69: 1278–85
13. Weller, T. H. 1971. The cytomegaloviruses: Ubiquitous agents with protein clinical manifestations (first of two parts). *N. Engl. J. Med.* 285: 203–14
14. Weller, T. H. 1971. The cytomegaloviruses: Ubiquitous agents with protein clinical manifestations (second of two parts). *N. Engl. J. Med.* 285:267–74
15. Fiala, M., Payne, J. E., Berne, T. V., Moore, T. L., Henle, W., Montgomerie, J. Z., Chatterjee, S. N., Guze, L. B. 1975. Epidemiology of cytomegalovirus infection after transplantation and immunosuppression. *J. Infect. Dis.* 132:421–33
16. Luby, J. P., Burnett, W., Hull, A. R., Ware, A. J., Shorey, J. W., Peters, P. C. 1974. Relationship between cytomegalovirus and hepatic function abnormalities in the period after renal transplant. *J. Infect. Dis.* 129:511–18
17. Prince, A. M., Brotman, B., Grady, G. F., Kahns, W. J., Hazzi, C., Levine, R. W., Millian, S. J. 1974. Long-incubation post-transfusion hepatitis without serological evidence of exposure to hepatitis B virus. *Lancet* 2:241–46
18. Prince, A. M., Szmuness, W., Millian, S. J., David, D. S. 1971. A serologic study of cytomegalovirus infections associated with blood transfusions. *N. Engl. J. Med.* 284: 1125–31
19. Stevens, D. P., Barker, L. F., Ketchem, A. S., Meyer, H. M.

1970. Asymptomatic cytomaglovirus infection following blood transfusion in tumor surgery. *J. Am. Med. Assoc.* 211:1341–44

20. Gregor, G. R., Gellar, S. A., Wolker, G. F., Compcmanes, B. A. 1975. Coxsackie hepatitis in an adult with ultrastructural demonstration of the virus. *Mt. Sinai J. Med. NY* 42:575–80
21. Purcell, R. H., Dienstag, J. L., Feinstone, S. M., Kapikian, A. Z. 1975. Relationship of hepatitis A antigen to viral hepatitis. *Am. J. Med. Sci.* 270:61–71
22. Dienstag, J. L., Davenport, F. M., McCollum, R. W., Hennessy, A. V., Klatskin, G., Purcell, R. H. 1976. Nonhuman primate-associated viral hepatitis type A. *J. Am. Med. Assoc.* 237:462–64
23. Szmuness, W., Dienstag, J. L., Stevens, L. E., Purcell, R. H. 1977. The prevalence of antibody to hepatitis A antigen in various parts of the world. *Am. J. Epidemiol.* In press.
24. Dienstag, J. L., Alaama, A., Mosley, J. W., Redeker, A. G., Purcell, R. H. 1977. Etiology of sporadic hepatitis B surface antigen-negative hepatitis. *Ann. Intern. Med.* 87:1–6
25. Villarejos, V. M., Visona, K. A., Eduarte A., C. A., Provost, P. J., Hilleman, M. R. 1975. Evidence for viral hepatitis other than type A or type B among persons in Costa Rica. *N. Engl. J. Med.* 293:1350–52
26. Galbraith, R. M., Portman, B., Eddleston, A. L., Williams, R. 1975. Chronic liver disease developing after outbreak of HBsAg-negative hepatitis in haemodialysis unit. *Lancet* 2:886–89
27. Muller, R., Stephan, B., Helmstedt, D. 1975. Occurrence of viral hepatitis type B and type non-B in three patients. *J. Infect. Dis.* 132: 195–99
28. Seeff, L. B., Wright, E. L., Zimmerman, H. J., McCollum, R. W., Members of the VA Hepatitis Cooperative Studies Group. 1975. VA cooperative study of posttransfusion hepatitis, 1969–1974: Incidence and characteristics of hepatitis and responsible risk factors. *Am. J. Med. Sci.* 270:355–62
29. Goldfield, M., Black, H. C., Bill, J., Srihongse, S., Pizzati, W. 1975. The consequence of administering blood pretested for HBsAg by third generation techniques. A progress report. *Am. J. Med. Sci.* 270:335–42
30. Knodell, R. G., Conrad, M. E., Ishak, K. G. 1977. Development of chronic liver disease after acute non-A, non-B post-transfusion hepatitis. *Gastroenterology* 72:902–9
31. Deinhardt, F., Holmes, A. W., Capps, R. B., Popper, H. 1967. Studies on the transmission of human viral hepatitis to marmoset monkeys. I. Transmission of disease, serial passages, and description of liver lesions. *J. Exp. Med.* 125:673–88
32. Parks, W. P., Melnick, J. L., 1969. Attempted isolation of hepatitis viruses in marmosets *J. Infect. Dis.* 120:539–47
33. Parks, W. P., Melnick, J. L., Voss, W. R., Singer, D. B., Rosenberg, H. S., Alcott, J., Lasazza, A. M. 1969. Characterization of marmoset hepatitis virus. *J. Infect. Dis.* 120: 548–59
34. Holmes, A. W., Deinhardt, F., Wolfe, L., Froesner, G., Peterson, D., Castro, B., Conrad, M. E. 1973. Specific neutralization of human hepatitis type A in marmoset monkeys. *Nature* 243:419–20
35. Mosely, J. W., Redeker, A. G., Feinstone, S. M., Purcell, R. H. 1977. Multiple hepatitis viruses in multiple attacks of acute viral hepatitis. *N. Engl. J. Med.* 296:75–78
36. Alter, H. J., Holland, P. V., Purcell, R. H. 1975. The emerging pattern of posttransfusion hepatitis. *Am. J. Med. Sci.* 270:329–34
37. Knodell, R. G., Conrad, M. E., Ginsberg, A. L., Bell, C. J., 1976. Efficacy of prophylactic gammaglobulin in preventing non-A, non-B post-transfusion hepatitis. *Lancet* 1:557–61
38. Miller, W. V., Watson, L. E., Holland, P. V., Walsh, J. H. 1971. Evaluation of the effectiveness of hepatitis screening tests. *Vox Sang.* 21:1–10
39. Molnar, I. G., Gitnick, G. L. 1977. *Gastroenterology* 72:A80/1103 (Abstr.)
40. Blumberg, B. S., Alter, H. J., Visnich, S. 1965. A "new" antigen in leukemia sera. *J. Am. Med. Assoc.* 191:541–46
41. WHO Expert Committee on Viral Hepatitis. 1977. Terminology of hepatitis viruses and antigens. *Intervirology* 8:65–67

Ann. Rev. Med. 1978. 29:367–78
Copyright © 1978 by Annual Reviews Inc. All rights reserved

HEPARIN, ANTITHROMBIN, AND ABNORMAL CLOTTING[1]

◆7286

Robert D. Rosenberg, M.D., Ph.D.[2]
Harvard Medical School, Sidney Farber Cancer Institute, Beth Israel Hospital,
Boston, Massachusetts 02115

INTRODUCTION

Thrombin is the end product of the coagulation cascade and a member of a class of proteins termed serine proteases (1). This type of enzyme possesses a unique serine moiety, which, in concert with residues in its immediate vicinity, enables the protein to cleave polypeptide bonds (2). Once thrombin is evolved within the blood, two phenomena can occur. On the one hand, the enzyme can scission two sets of arginine-glycine bonds in fibrinogen. This action releases fibrinopeptides A and B from the macromolecule and thus converts it to fibrin (3). Fibrin can then undergo a complex set of polymerization reactions that result in the formation of the thrombus. On the other hand, thrombin can be inhibited within the blood, and fibrin formation can thereby be prevented. On the basis of this latter set of reactions, it has long been suspected that a specific inactivator of thrombin must be present under normal physiologic conditions (4–7). This inhibitor has been termed antithrombin. Recently, this component has emerged as a critical control point in the coagulation mechanism and a major barrier against thrombus formation (8–12). Furthermore, it has become apparent that the function of heparin as an anticoagulant is critically dependent upon the presence of this macromolecule. In this review, we briefly summarize the current status of knowledge in this area. In addition, we indicate its implications with respect to the pathophysiology of thrombosis, as well as the pharmacology of heparin's anticoagulant action.

MOLECULAR CHARACTERISTICS OF ANTITHROMBIN AND HEPARIN

Antithrombin

Human antithrombin has been isolated from plasma in purified form by a variety of procedures (8). The final product that we utilized is physically homogeneous,

---

[1] Supported by the National Institutes of Health grant #HL19131
[2] Recipient of American Heart Association Established Investigatorship Award

0066-4219/78/0401-0367$01.00

as judged by various immunoelectrophoretic and electrophoretic techniques (8). To demonstrate that this single protein possesses both antithrombin and heparin cofactor activity, we examined the inactivation of purified human thrombin in the presence and absence of heparin (Figure 1). If thrombin is added to buffer with or without heparin, no loss in enzymatic activity is noted. If thrombin is incubated for varying periods of time with antithrombin, a slow, progressive decline in the enzyme's activity is observed ("antithrombin effect"). If thrombin is incubated with a mixture of heparin and antithrombin, its enzymatic activity is inhibited virtually instantaneously ("heparin-cofactor effect"). To prove that this unique macromolecule is responsible for most of the antithrombin and heparin cofactor activities normally present in plasma, we immunoprecipitated this component from defibrinated plasma with γ-globulin directed against the inhibitor and measured both of these residual activities. The data revealed that as the concentration of γ-globulin is increased both antithrombin and heparin cofactor activities decline in parallel fashion, and that less than 10% of either activity remains at the highest level of γ-globulin utilized (8).

Heparin

Heparin is a mucopolysaccharide of molecular weight 5000–50,000. This natural product is found in a variety of organs including the heart, the liver, the intestine, etc. The ultimate structure of the mucopolysaccharide is depicted in Figure 2. However, it is important to note that the mucopolysaccharide sequence shown represents only one of many possible arrangements of hexosamine and uronic acid residues. Furthermore, it is known that heparin preparations exhibit considerable polydispersity in molecular size (13), variation in the ratio of glucuronic acid to iduronic acid (14), alteration in the amount of ester and *N*-sulfation (15), as well as differing extents of *N*-acetylation (16). Changes in each of these parameters correlate only to a limited degree with heparin's anticoagulant potency. Thus the precise relationship between the structure of heparin and its anticoagulant properties remains elusive. Indeed, it has been tacitly assumed that this mucopolysaccharide may exhibit a variety of alternative structures each equally capable of interacting with antithrombin.

We thought this assumption unlikely to be true, and considered the alternative possibility that heparin preparations were grossly impure and that only a small fraction of the molecules were responsible for the distinctive anticoagulant properties. In this case, attempts to relate the structure of the bulk material to anticoagulant potency would be futile. Numerous investigators have fractionated heparin preparations by gel filtration, ion-exchange chromatography, or electrophoresis without observing any substantial increase in purity (17, 18). Therefore, we decided to develop a biologic approach to fractionation and attempted to employ heparin's specific affinity for antithrombin as the means for obtaining highly active preparations of the mucopolysaccharide (19). With this technique we were able to show that all standard heparin preparations consist of two distinct forms that differ greatly in their ability to interact with antithrombin. The first form constitutes approximately two thirds of the chemical mass of the

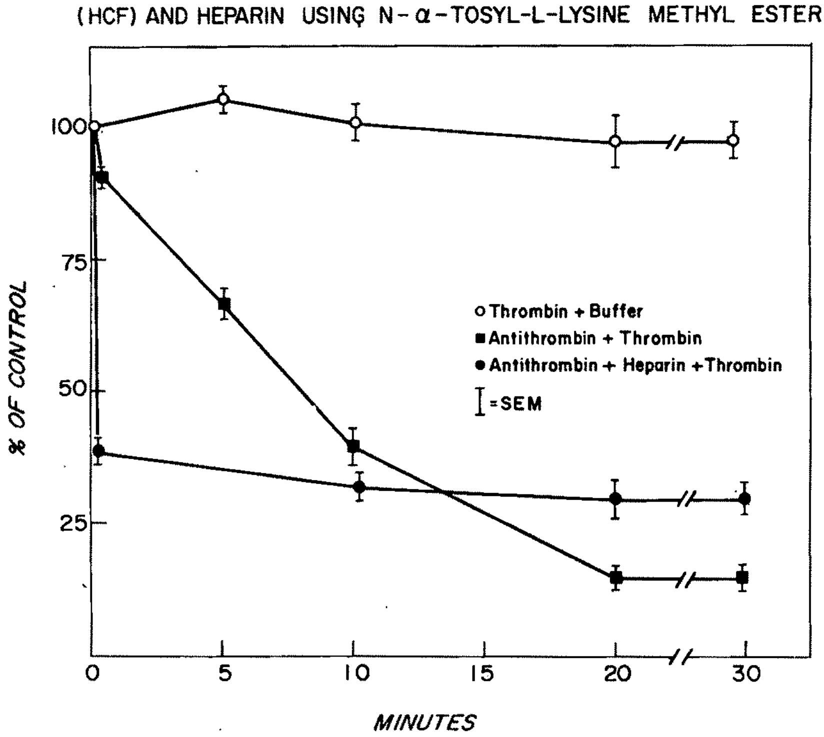


Figure 1 Inhibition of thrombin by antithrombin-heparin cofactor in the presence and absence of heparin. Final concentrations of enzyme and inhibitor were 0.162 and 0.164 absorbance units per ml, respectively. Final concentration of added heparin was 12 units/ml. Esterolytic activity of thrombin was measured with tosyl-L-lysine methyl ester (Cyclo Chemical Co., Los Angeles) by continuously recording uptake of 0.01 *M* NaOH in a pH-stat fitted with automatic burette and chart recorder. Buffer was 0.005 *M* Tris-HCl in 0.15 *M* NaCl (pH 8.3). All protein preparations were extensively dialyzed against this assay buffer. Concentration of substrate was 0.015 *M*, and reaction velocity at 37°C ± 0.1°C was recorded for the first 3 min. Plots of thrombin concentration vs reaction velocity were linear, with remarkably constant slope and intercept values.

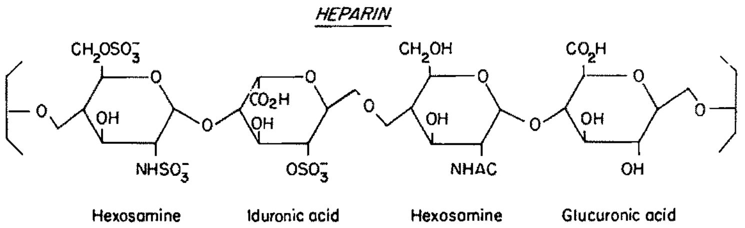


Figure 2 Structure of heparin.

unfractionated heparin and does not bind to antithrombin under the conditions of these experiments. This fraction is responsible for only about 15% of the total anticoagulant activity of the original material. The second form, comprising one third of the remaining chemical mass of unfractionated heparin, binds tightly to antithrombin and contains approximately 85% of the total anticoagulant activity of the original material. The presence of these two distinct forms of the mucopolysaccharide has been confirmed by work from another laboratory (20). However, we have recently found that the active form of heparin is itself composed of a family of molecular species with varying anticoagulant function (21). Indeed, careful fractionation of this active pool has yielded macromolecules with potencies that are ten times that of the standard heparin preparations. This remarkable heterogeneity is an intrinsic property of the mucopolysaccharide and is not a result of the commercial processing of porcine intestinal mucosa or beef lung, the primary sources of heparin. At the present time, we have only incomplete knowledge of the specific features that distinguish the biologically active from the relatively inactive heparin species. However, work from our laboratory indicates that a specific sequence of uronic acid and glucosamine residues is essential. In particular, the precise placement of glucuronic acid residues appears to be a prerequisite for anticoagulant function (21).

MECHANISM OF ANTICOAGULANT ACTION

We studied the neutralization of thrombin by antithrombin and demonstrated that the equivalence point of the reaction occurs at a 1:1 stoichiometric ratio of enzyme to inhibitor. The addition of heparin to these components does not appreciably alter the equivalence point, but as shown above, does dramatically accelerate the rate of the reaction (8) (Figure 1).

We also directly monitored the physical formation of a remarkably stable thrombin-antithrombin complex. As expected from estimates of reaction stoichiometry, the complex has a molecular weight approximately equal to the sum of the molecular weights of thrombin and antithrombin. In the absence of heparin, complex formation occurs at a relatively slow rate, consonant with the gradual inhibition of enzymatic activity. In the presence of heparin, complex formation is virtually instantaneous (8).

The formation of this enzyme-inhibitor complex, both in the presence and absence of heparin, is critically dependent upon the active serine center of thrombin. If this residue is blocked by the addition of diisopropylfluorophosphate, interaction between thrombin and antithrombin is totally inhibited. As previously noted, thrombin's capacity to transform fibrinogen to fibrin depends on its ability to specifically cleave arginine bonds. Therefore we expected that the reactive site of the inhibitor that binds the active serine center of the enzyme would contain an arginine residue. Indeed, our modification of this group on antithrombin with 2,3-butanedione or 1,2-cyclohexanedione virtually eliminates the ability of this protein to inhibit thrombin in the presence or absence of heparin (8).

Given the highly acidic nature of heparin, one would expect that positive groups on antithrombin such as ϵ-amino lysyl residues form the binding site for this negatively charged anticoagulant. We have shown that the chemical alteration of these residues with *o*-methylisourea prevents the binding of heparin to this protein and suppresses the acceleration of inhibitor action by this anticoagulant. However the slow, progressive neutralization of thrombin by antithrombin is not appreciably affected (8).

Furthermore, we recently demonstrated that the active form of heparin functions as a catalyst in this reaction. Relatively small amounts of this mucopolysaccharide are able to dramatically accelerate the interaction of considerably larger amounts of thrombin and antithrombin (22). The ability of heparin to function in this fashion would be explicable if it were to be displaced from antithrombin during formation of the thrombin-antithrombin complex. Thus, the mucopolysaccharide would be available to bind to free inhibitor and cyclically promote subsequent rounds of interactions. To establish the validity of this mechanism we sought direct evidence for the displacement of mucopolysaccharide (22). This was accomplished with isotopically labelled heparin in conjunction with a technique for separating free mucopolysaccharide from heparin bound to inhibitor. The addition of thrombin to the preformed heparin-antithrombin complex results in the release of mucopolysaccharide. The displacement of heparin is proportional to the amount of thrombin added. Complete liberation of mucopolysaccharide occurs when the molar concentration of added thrombin is equal to the molar concentration of the heparin-antithrombin complex (22).

On the basis of the above data, it would appear that antithrombin neutralizes the activity of thrombin by complex formation via a reactive site (arginine)–active center (serine) interaction. If small amounts of heparin are added to the system, it preferentially binds to the lysyl residues on antithrombin. The resulting heparin-antithrombin complex rapidly inactivates thrombin. This is most probably due to a heparin-dependent conformational alteration of the inhibitor, which renders the reactive site arginine more accessible to the active serine center of thrombin. Once thrombin-antithrombin complex formation has occurred, heparin is released and is again available for binding to free inhibitor. Thus the mucopolysaccharide is capable of catalyzing numerous subsequent rounds of thrombin-antithrombin complex formation (Figure 3).

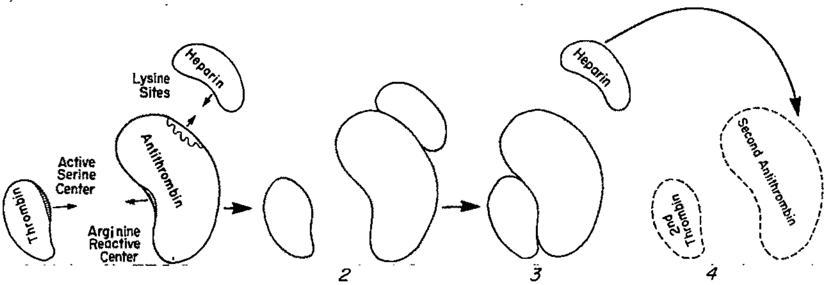


Figure 3 Mechanism of heparin action.

The coagulation cascade is composed of a series of linked proteolytic reactions. At each stage of this mechanism, a parent zymogen is converted to a corresponding serine protease, which catalyzes a subsequent zymogen-serine protease transition. Thus, of the eight proteins directly or indirectly involved in the conversion of prothrombin to thrombin, five proteins are ultimately activated to serine proteases (Factors XIIa, XIa, IXa, VIIa, and Xa), and three proteins are cofactors for these proteolytic events (high-molecular-weight kininogen, and Factors VIII and V). The fibrinolytic system is similar in design to the coagulation mechanism. Furthermore, its end product is the serine protease plasmin.

Previous investigators have shown that Factor Xa is neutralized by antithrombin and heparin (23–25). On the basis of this evidence and our knowledge of the biochemical mechanism of antithrombin, we predicted that all serine proteases produced within the coagulation and fibrinolytic systems would be neutralized by this inhibitor in similar fashion, and that heparin would accelerate each of these interactions (9, 26).

To test this concept, we examined the interaction of antithrombin and heparin with these serine proteases: Factor IXa (10), Factor XIa (9), Factor XIIa (11, 12), and plasmin (27). In each of these cases, a slow but progressive inhibition of enzymatic activity occurred in the presence of antithrombin. This process of inactivation is due to the formation of a 1:1 stoichiometric complex of enzyme and inhibitor. In each instance, the addition of heparin dramatically accelerates the rate of the neutralization without altering the stoichiometry or nature of complex formation (Figure 4). These unsuspected observations have recently been confirmed by work from other laboratories (28–30). The only possible exception to these findings may be in the inhibition of Factor VIIa by antithrombin. There has been a preliminary report that heparin's effect on this interaction is minimal (31).

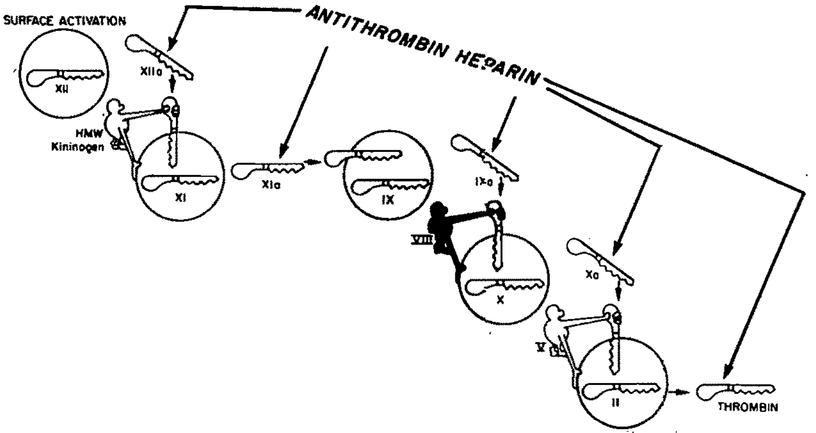


Figure 4 The sites of action of antithrombin and heparin within the coagulation cascade.

PHARMACOLOGIC IMPLICATIONS

Possible Side Effects of Commercial Preparations of Heparin

As we noted in previous sections of this review, heparin preparations presently in clinical usage are quite heterogeneous. Indeed, only 25–35% of the bulk mucopolysaccharide has any appreciable ability to bind to and activate antithrombin. The "inactive" mucopolysaccharide species present in all heparin preparations may be responsible for undesirable side effects of this medication. It has been known for some time that heparin can induce platelets to undergo release reactions in vitro (32). Recently, clinical studies have revealed that substantial numbers of patients who are being given this drug experience significant reductions in their platelet count (33). These observations suggest that heparin infusions may activate platelets, cause their deposition within the vasculature, and thus produce systemic thrombocytopenia. If such is the case, perhaps heparin species incapable of binding to antithrombin are available in plasma for interactions with platelets and are responsible for this unwanted phenomenon.

Clinical Assay of Heparin Function

Heparin can be employed in certain groups of presurgical patients for prophylaxis against the development of venous thrombosis and its sequela (34). In addition, this drug can be utilized to treat established venous thrombotic disease and dramatically reduce the incidence of pulmonary embolism (35). In order to extend the prophylactic usage of this anticoagulant to other groups of patients and improve the therapy of established thrombotic disease, we must consider the inadequacies of the present assay utilized for monitoring drug administration, as well as the identification of clinical states that may be refractory to heparinization.

From the above review of heparin's function, it should be obvious that optimal dosage of this anticoagulant is dependent upon a knowledge of the concentrations and relative distributions of the serine proteases that this mucopolysaccharide is instrumental in neutralizing. Thus, the dosage of heparin required for an individual patient should be adjusted according to the procoagulant stimulus with which this drug must deal. Therefore, a patient with activation of the early-clotting factors (Factors XIIa and XIa), whose concentrations in plasma are low, may require only a minimal dosage of heparin in order to assure prophylaxis against thrombosis. On the other hand, an individual with activation of the clotting mechanism so extensive as to have produced Factor Xa or thrombin, which have much larger concentrations in plasma, may require a significantly greater dosage to achieve the same degree of protection.

Present assays attempt to regulate heparin therapy by determining the degree to which plasma from a heparinized patient inhibits the in vitro addition of an arbitrary amount of either thrombin or Factor Xa, or else the extent to which this plasma prevents in vitro generation of an arbitrary amount of serine proteases during kaolin activation (activated partial thromboplastin time)(36–38). In either case, no attempt is made to judge the effectiveness of the heparin-activated plasma antithrombin in inhibiting the serine proteases of the coagulation cascade generated within the patient's circulatory system. In effect, it is as if one were to treat a diabetic patient by measuring the amount of injected insulin present in his circulatory system rather than the resultant reduction in blood glucose levels. This critique of present methodology is pertinent to situations in which either standard dosages of heparin are employed after thrombus formation, or low dosages are administered prior to clot development.

Several groups of investigators are establishing methods for the direct or indirect measurement of in vivo levels of the serine proteases active in coagulation. With the advent of more sophisticated assay procedures, it may become apparent that certain patients are refractory to heparinization. For example, Nossel et al (39) measured fibrinopeptide A levels in individuals with pulmonary embolism who were treated with this drug. In most of these patients, heparin administration results in an immediate reduction of fibrinopeptide A levels. This is consistent with the half-life of this small molecular species and suggests that thrombin's action has been instantaneously suppressed. In some individuals, however, fibrinopeptide A evolution remained elevated during the first 30 minutes after injection of 10,000 units of heparin. We interpret these data to indicate that this group of patients are resistant to heparin therapy, despite the fact that samples of their blood are incoagulable, as judged by clinical assay procedures (partial-thromboplastin time, whole-blood clotting time, etc.).

Based upon our knowledge of the biochemistry of heparin, one can suggest a variety of molecular explanations for these refractory states. The simplest mechanisms would involve a more rapid clearance of heparin, grossly reduced concentrations of antithrombin, or an increase in the levels of the molecular species other than antithrombin that interact with heparin. These phenomena, however, are unlikely to play a significant role in the genesis of refractory states considered above, since the blood must have been grossly incoagulable immediately after drug administration. The more likely explanation is that refractory states are due to various local events that might permit in vivo clotting, but whose effects are not measurable by currently employed assay techniques. This paradoxical state could be caused by the sequestration of thrombin such that it is inaccessible to heparin action, or by localized reductions in the level of anticoagulant. For example, we demonstrated that a serine protease similar to activated intermediates of the coagulation cascade is present on the macrophage membrane (40). This enzyme is neutralized by antithrombin, but when heparin is added to the inhibitor, the inactivation is prevented. This phenomenon is due to the exclusion of the negatively charged heparin-antithrombin complex from the similarly charged macrophage surface. In an analogous manner, thrombin may be adsorbed to damaged endothelium or platelet membrane in such a way as to render it inaccessible to heparin-antithrombin neutralization yet capable of local thrombus formation. An alternate mechanism for generating a refractory state would involve a decrease in the local availability of heparin. This could occur by local release of large quantities of platelet factor 4, which directly neutralizes heparin. This phenomenon could be due to altered local blood flow properties or restricted surface abnormalities that induce platelet release reactions. The local anticoagulant function of heparin might be exhausted and fibrin deposition could occur, as evidenced by the fibrinopeptide A release. The systemic levels of this drug as judged by the activated partial thromboplastin time, the whole-blood clotting time, and the Factor Xa–inhibition assays would be unaffected.

PHYSIOLOGIC IMPLICATIONS

This newly acquired knowledge of the structure and function of heparin may have important physiologic implications. The relation of in vitro inhibitor-heparin interactions to in vivo modulation of the coagulation system remains speculative. The data outlined above demonstrate that this acidic mucopolysaccharide, in conjunction with antithrombin, inhibits in an instantaneous fashion virtually all of the enzymatic steps of the coagulation system. This is to be contrasted with the relatively slow neutralization of many of these enzymes by other circulating plasma inhibitors.

To examine the possible relevance of this inhibitory mechanism to normal physiological processes, let us consider the sequela of congenital antithrombin deficiency vis-à-vis the plasma levels of this inhibitor, and examine the probable in vivo location of heparin-like material.

Congenital reductions in antithrombin levels are uniquely associated with fatal thrombotic complications. The first report of this syndrome in 1965 described several generations of a family with recurrent venous thromboembolism in conjunction with a plasma antithrombin concentration that averaged 40% of normal (41). Other families were subsequently reported to exhibit venous and arterial thrombotic episodes in association with partial deficiencies of antithrombin (42, 43).

The data presented earlier indicate that the in vitro addition of antithrombin to serine proteases of the coagulation mechanism results in the neutralization of their proteolytic activity only after many minutes have passed. Thus, the production of potent serine proteases should trigger clot development long before inhibition by antithrombin has occurred, and it is surprising that modest reductions in plasma concentrations of this molecular species result in such striking thrombotic phenomena.

The discrepancy between clinical observations of inhibitor deficiency and in vitro kinetics of serine protease–antithrombin interaction can be resolved, if heparin-like material were shown to activate this component in vivo. Under these conditions, plasma antithrombin would almost instantaneously inactivate serine proteases, and a reduction in its concentration would have profound pathophysiological effects.

Heparin has not only been isolated from various organs but is also thought to be present in most cells. More importantly, heparan sulfate, which possesses some anticoagulant properties, has been found on a variety of cell surfaces including those of the endothelium (44) and the platelet (45). How much of the "highly active" hemostatically potent fraction of heparin is present at these locales remains to be determined. However, the availability of heparin-like components would permit antithrombin to be selectively activated at blood-surface interfaces where enzymes of the hemostatic mechanism are generated. Thus, the plasma protease inhibitor would be critically placed to neutralize these enzymes and thereby protect natural surfaces against thrombus formation.

Furthermore, the catalytic nature of heparin would ensure the continual regeneration of the nonthrombogenic properties of these natural surfaces. Once the antithrombin bound to platelet surface or vessel wall mucopolysaccharide has complexed with enzyme, the enzyme-inhibitor complex would be liberated into the circulation. The heparin-like material would again be available to recruit free antithrombin and thereby continually renew the ability of the surface to resist the attack of serine proteases of the hemostatic mechanism. Alterations of this protective barrier could be responsible for early arterial or venous prethrombotic lesions in man.

Literature Cited

1. Reich, E., Rifkin, D., Shaw, E., eds. *Proteases and Biological Control*. 1975. New York: Cold Spring Harbor Laboratory. 1021 pp.
2. Gladner, J. A., Laki, K. 1958. The active site of thrombin. *J. Am. Chem. Soc.* 80:1263–64
3. Blomback, B., Blomback, M., Hessel, B., Iwanaga, S. 1967. Structure of *N*-terminal fragments of fibrinogen and specificity of thrombin. *Nature* 215:1445–48
4. Contejean, C. 1895. Recherches sur les injections intraveineuses de peptone et leur influence sur la coagulabilite du sang chez le chien. *Arch. Physiol. Norm. Pathol.* 7:45–53
5. Brinkhous, K. M., Smith, H. P., Warner, E. D., et al. 1939. The inhibition of blood clotting: An unidentified substance which acts in conjunction with heparin to prevent the conversion of prothrombin into thrombin. *Am. J. Physiol.* 125:683–87
6. Waugh, D. F., Fitgerald, M. A. 1956. Quantitative aspects of antithrombin and heparin in plasma. *Am. J. Physiol.* 184:627–39
7. Abildgaard, U. 1968. Highly purified antithrombin III with heparin cofactor activity prepared by disc electrophoresis. *Scand. J. Clin. Lab. Invest.* 21:89–91
8. Rosenberg, R. D., Damus, P. S. 1973. Purification and mechanism of action of human antithrombin-heparin cofactor. *J. Biol. Chem.* 248: 6490–6505
9. Damus, P. S., Hicks, M., Rosenberg, R. D. 1973. A generalized view of heparin's anticoagulant action. *Nature* 246:355–57
10. Rosenberg, J. S., McKenna, P., Rosenberg, R. D. 1975. Inhibition of human Factor IXa by human antithrombin-heparin cofactor. *J. Biol. Chem.* 250:8883–88
11. Stead, N., Kaplan, A., Rosenberg, R. D. 1976. Inhibition of Factor XIIa with antithrombin. *Clin. Res.* 24:321A
12. Stead, N., Kaplan, A. P., Rosenberg, R. D. 1976. Inhibition of activated Factor XII by antithrombin-heparin cofactor. *J. Biol. Chem.* 251:6481–88
13. Laurent, T. C. 1961. Studies on fractionated heparin. *Arch. Biochem. Biophys.* 92:224–31
14. Hovingh, P., Linker, A. 1970. Enzymatic degradation of heparin and heparitin sulfate. *J. Biol. Chem.* 245:6170–75
15. Danishefsky, I., Steiner, H., Bella, A., Friedlander, A. 1969. Investigations on the chemistry of heparin. *J. Biol. Chem.* 244:1741–45
16. Cifonelli, J. A., King, J. 1973. Structural studies on heparins with unusually high *N*-acetylglucosamine contents. *Biochim. Biophys. Acta* 320:331–41
17. Ehrlich, J., Stivala, S. S. 1973. Chemistry and pharmacology of heparin. *J. Pharm. Sci.* 62:517–44
18. Cleland, R. C., Cleland, M. C., Lipsky, J. J. 1968. Ionic poly-

saccharides I. Adsorption and fractionation of polyelectrolytes on (diethylamino)ethyl cellulose. *J. Am. Chem. Soc.* 90:3141–46

19. Lam, L. H., Silbert, J. E., Rosenberg, R. D. 1976. Separation of active and inactive forms of heparin. *Biochem. Biophys. Res. Commun.* 69:570–77
20. Höök, M., Bjork, I., Hopwood, J., Lindahl, U. 1976. Anticoagulant activity of heparin: Separation of high-activity and low-activity heparin species by affinity chromatography on immobilized antithrombin. FEBS Lett. 66:90–93
21. Rosenberg, R. D., Armand, G., Lam, L. H. Heparin-antithrombin interactions, *Proc. Int. Symp. Heparin, Saskatoon, Saskatchewan Meet.* New York: Academic. In press
22. Rosenberg, R. D. 1977. Biological actions of heparin. In *Seminars In Hematology,* Vol. XIV, pp. 427–40. New York: Grune & Stratton
23. Seegers, W. H., Cole, E. R., Harmision, C. R., et al. 1964. Neutralization of autoprothrombin C activity with antithrombin. *Can. J. Biochem.* 42:359–64
24. Biggs, R., Denson, K. W. E., Akman, N., Barrett, R., Hadden, M. 1970. Antithrombin III, antifactor Xa and heparin. *Br. J. Haematol.* 19:283–305
25. Yin, E. T., Wessler, S., Stoll, P. J. 1971. Rabbit plasma inhibitor of the activated species of blood coagulation Factor X: Purification and some properties *J. Biol. Chem.* 246:3694–3702
26. Rosenberg, R. D. 1973. Heparin action, *Circulation* 49:603–5
27. Highsmith, R. F., Rosenberg, R. D. 1974. Inhibition of human plasmin by human antithrombin-heparin cofactor. *J. Biol. Chem.* 249:4335–38
28. Kurachi, K., Fujikawa, K., Schmier, G., Davie, E. W. 1976. Inhibition of bovine Factor IXa and Factor Xaβ by antithrombin III. *Biochemistry* 15:373–77
29. Chan, J., Burrowes, C., Habal, F., Movat, H. 1977. Inhibition of activated Factor XII (Hageman Factor) by antithrombin III: The effect of other plasma proteinase inhibitors. *Biochem. Biophys. Res. Commun.* 74:150–58
30. Summaria, L., Boreisha, I. G., Arzadon, L., Robbins, K. 1977. Activation of human glu-plasminogen to glu-plasmin by urokinase in presence of plasmin inhibitors. Streptomyces leuteptin & human plasma alpha-1-antitrypsin & antithrombin III. *J. Biol. Chem.* 252:3945
31. Godal, H. C., Ryah, M., Laake, K. 1974. Progressive inactivation of purified Factor VII by heparin and antithrombin III. *Thromb. Res.* 5:773–76
32. Eika, C. 1972. On the mechanism of platelet aggregation induced by heparin and polybrene. *Scand. J. Haematol.* 9:248–57
33. Bell, W. R. 1976. Thrombocytopenia occurring during heparin therapy. *N. Engl. J. Med.* 295:276–77
34. Kakkar, V. V., Nicolaides, A. N., Field, E. S., Flute, P. T. 1971. Low doses of heparin in prevention of deep-vein thrombosis. *Lancet* 2:669–77
35. Hirsh, J., Genton, E. 1975. Low-dose heparin prophylaxis for venous thromboembolism. In *Prophylactic Therapy of Deep Vein Thrombosis and Pulmonary Embolism* J. Frantantoni, S. Wessler, eds., pp. 183–206. Washington, DC: DHEW Pub. No. (NIH) 76–866
36. Blomback, B., Blomback, M., Olsson, P., et al. 1959. Determination of heparin level of the blood: Some observations on heparin elimination and correlation between heparin level and clotting time after intravenous injection. *Acta Chir. Scand. Suppl.* 245:259–64
37. Yin, E. T., Wessler, S., Butler, J. V. 1973. Plasma heparin: A unique, practical submicrogram-sensitive assay. *J. Lab. Clin. Med.* 81:298–310
38. Marder, V. J. 1970. A simple technique for the measurement of plasma heparin concentration during anticoagulant therapy. *Thromb. Diath. Haemorrh.* 24:230–39
39. Nossel, H. L., Younger, L. R., Wilner, G. D., Procupez, T., Canfield, R. D., Butler, V. P. 1971. A radioimmunoassay for human fibrinopeptide A. *Proc. Natl. Acad. Sci.* 68:2350–53
40. Remold, H. G., Rosenberg, R. D 1975. Enhancement of migration inhibitory factor (MIF) by plasma esterase inhibitor. *J. Biol. Chem.* 250:6608–13
41. Egeberg, O. 1965. Inherited anti-

thrombin deficiency causing thrombophilia. *Thromb. Diath. Haemorrh.* 13:516–30

42. Shapiro, S. S., Prager, D., Martinez, J. 1973. *Inherited antithrombin III deficiency associated with multiple thromboembolic phenomena.* Presented at the Ann. Meet. Am. Soc. Hematol., 16th, Chicago, p. 63
43. Grunberg, J., Smallridge, R., Rosenberg, R. D. 1975. Inherited antithrombin III deficiency causing mesenteric venous infarction. *Ann. Surg.* 181:791–94
44. Buonassisi, V. 1973. Sulfated mucopolysaccharide synthesis and secretion in endothelial cell cultures. *Exp. Cell Res.* 76:363–68
45. Horner, A. A. 1975. Demonstration of endogeneous heparin in rat blood. In *Heparin: Structure, Function and Clinical Implications.* ed. R. A. Bradshaw, S. Wessler, p. 85–91 New York: Plenum

Ann. Rev. Med. 1978. 29:379–406
Copyright © 1978 by Annual Reviews Inc. All rights reserved

HUMORAL SYNDROMES ASSOCIATED WITH CANCER

◆7287

W. D. Odell, M.D., Ph.D and A. R. Wolfsen, M.D.
Departments of Medicine and Physiology, Harbor General Hospital, UCLA School of Medicine, Torrance, California 90509

INTRODUCTION

In addition to symptoms produced by direct effects of tumor mass, neoplasms may produce symptoms by elaboration of humoral substances. Heightened physician awareness of these syndromes in recent years has led to increasing frequency of diagnosis, increasing cognizance of the kinds of humoral substances produced, and the use of such humoral substances as tumor markers to assist in diagnosis and in following response to therapy. As endocrinologists, our interests relate not only to the clinical expression of such humoral production but to a fundamental understanding of the structure, measurement, and biochemical actions of these humoral substances. Some years ago, we analyzed (*a*) the spectrum of the clinically recognizable humoral syndromes associated with cancer, (*b*) the fact that where known (with a single exception) the syndromes are caused by elaboration of peptides or proteins, and (*c*) the further fact that elaboration of biologically inactive peptides is more common than elaboration of biologically active peptides (1, 2). As a result we postulated that ectopic peptide elaboration is a universal concomitant of neoplasia. Additional data collected during the past one to two years have further substantiated this hypothesis (3).

The many varied morphological and functional cell types of a multicellular organism contain identical data coded into the nuclear nucleic acid content. It is estimated that over 90% of this information is repressed or inactivated in the differentiated cell. Any or all such data may be "derepressed" or activated during cell replication. Such activation results in new protein synthesis. In addition, it is possible that genic data is never fully repressed or inactivated, but that in all cells variable quantities of all genic data are in fact translated into small amounts of protein synthesis. The control mechanisms for inactivation-activation are not fully understood, but in addition to simple cell replication modifying repression, the process of neoplasia per se may modify repression controls. The distinction between ectopic peptide elaboration by replicating

0066-4219/78/0401-0379$01.00

cells versus ectopic peptide elaboration by neoplastic cells is obviously of great importance. If these peptides are to be useful as tumor markers they must be relatively (quantitatively) or absolutely specific for the neoplastic process.

Current data (e.g. 4–14) lead one to speculate that all peptide hormones (and possibly a variety of other peptides) evolve during synthesis, storage, secretion, and degradation, through what we term the Peptide Cascade:

$$\begin{tikzcd}
 & & & \text{CARBOXYL FRAGMENT} \\
\text{PREPROHORMONE} \arrow[r] & \text{PROHORMONE} \arrow[r] & \text{HORMONE} \arrow[ur] \arrow[dr] & \\
 & & & \text{AMINO FRAGMENT}
\end{tikzcd}$$

The hormone begins in the prepro form, which contains methionines, the universal codon for protein synthesis, plus a variable-length additional amino acid sequence of unknown or poorly understood significance. This additional sequence may be important in intracellular channeling to storage sites. The preprohormone is degraded to the prohormone, which generally is biologically inactive or poorly active[1] (as is the preprohormone). This is in turn enzymatically degraded to the bioactive hormone. Either the bioactive hormone or the prohormone may be stored, depending on the particular hormone; cleavage may occur as part of the secretory process into blood. Once in blood, the hormone is degraded in peripheral tissues or by circulating peptidases into fragments; the carboxyl fragment is generally biologically inert, the amino fragment may be biologically active.

Current data reveal that biologically inactive fragments or precursor forms of hormones may commonly be elaborated by neoplasms. Some of these biologically inactive forms are probably prohormones.

In addition to the peptide cascade and its relation to neoplasia, additional biochemical data are pertinent and relate to the glycopeptide hormones: thyrotropin (TSH), luteinizing hormone (LH), follicle-stimulating hormone (FSH), and chorionic gonadotropin (hCG). This is presented only briefly here and is dealt with more fully in the section on gonadotropin elaboration. LH, FSH, TSH, and hCG are composed of two peptide chains, an alpha and a beta chain. The alpha chain is identical for LH, FSH, and TSH and differs for hCG by only a two-residue amino acid inversion and a three-residue deletion at the NH_2 terminus (15). The beta chain of each of these hormones is biochemically unique, conferring immunological and biological specificity (15–23). Neither the alpha nor the beta chain has any biological potency per se, but the alpha chain combined with a beta chain produces a biologically active molecule whose bioactivity relates to the beta chain. Cancers have been shown to elaborate the alpha chain, the beta chain, and/or intact hCG-like gonadotropin.

In summary, a variety of biologically inert-peptide hormone precursors, fragments, or one of the chains of the glycopeptides may be elaborated by cancers. While these may be excellent tumor markers, they produce no clinically recognizable syndromes. Elaboration of biologically active peptides occurs even less frequently than elaboration of biologically inert peptides. However, elaboration by neoplasms of either bioactive or inactive peptides is extremely common and may be a universal property of neoplasms.

---

[1] The exception is progastrin, which is biologically active.

HYPERCALCEMIA

Hypercalcemia is a relatively common finding in patients with carcinoma. Warwick et al (24) in 1961 reported a survey of 438 patients admitted to Toronto General Hospital for radio-therapy of cancer; of these, 9.1% had hypercalcemia. Laird-Meyers (25) in 1960 studied 430 patients with cancer-associated hypercalcemia; 56 of these hypercalcemic patients did not have demonstrable bony metastases. The most common carcinoma associated with hypercalcemia is that of the breast, and usually bony metastases are present. However, most kinds of carcinomas have been reported to be associated with hypercalcemia, particularly carcinomas of the lung, kidney, and cervix (1, 24, 25). In a prospective study of 200 patients with bronchogenic carcinoma, Bender & Hansen (26) found 12.5% (25 patients) had hypercalcemia. The incidence varied with histological type: 23% of patients with epidermoid carcinoma; 12.7% with large-cell anaplastic carcinoma; 2.5% with adenocarcinoma, and none with small-cell carcinoma. Fourteen of the 25 patients (56%) with elevated calcium did not have bony metastases.

The possible cause of this hypercalcemia in some patients with cancer was first suggested in 1941 by Albright (27), who discussed a patient with renal-cell carcinoma and metastases to the ileum. This patient had hypercalcemia that was attributed to the skeletal metastases. Albright pointed out that if this were the cause, both calcium and phosphorus should be elevated, but this man had a low blood phosphorus. Furthermore, parathyroid gland exploration during life and at autopsy showed no abnormalities. Albright thus suggested that the cancer elaborated a substance with parathormone-like actions. In 1964, Tashjian et al (28) first offered some substantiation of this hypothesis by identifying, through use of immunochemical techniques, a parathormone-like substance in carcinomas associated with hypercalcemia. Subsequently, similar identification was reported by Berson & Yalow (29) and Sherwood et al (30).

Buckle et al (31) studied a single patient with adenocarcinoma of the kidney in great detail. Extracts of tumor tissue gave a dose-response curve identical to bovine reference preparation; the tumor contained 2.2 μg/ml of parathyroid hormone–like activity. Most convincing, the renal artery and vein that supplied the kidney with the tumor contained 2.55 and 0.62 ng/ml parathormone activity, representing an arteriovenous difference of 1.93 ng/ml. Postoperatively, plasma parathormone fell to normal.

From these data and the clinical and chemical similarities of the hypercalcemia of cancer to hyperparathyroidism, it was believed until recently (1, 32, 33) that all or most patients with this syndrome, but without osseous metastases, had carcinoma elaboration of parathormone. However, as parathormone assays became more available, patients who had undetectable plasma parathormone were encountered. For example, in 1971 Federman, Sherwood & Potts (34) discussed a patient with squamous-cell carcinoma of the lung without osseous metastases, and with hypercalcemia and hypophosphatemia, and no detectable plasma parathormone. In 1973, Powell et al (35) examined four patients with bronchogenic carcinoma and hypercalcemia and found no parathormone activity either in their blood or in extracts of tumor tissue.

Another substance capable of producing such hypercalcemia has been described by Tashjian et al. In 1972 these workers (36) reported that the bone resorption–stimulating factor produced by fibrosarcoma in mice was a prostaglandin (PG), probably PGE_2. The hypercalcemia was treatable by indomethacin, a potent inhibitor of PG synthesis. Similar findings were demonstrated in a rabbit model, the VX_2 carcinoma (37). Prostaglandins had previously been shown to be potent stimulators of bone resorption in vitro (38, 39). Brereton et al (40) described a patient with adenocarcinoma of the kidney, without osseous metastases but with hypercalcemia, who also responded with a dramatic fall of blood calcium to normal upon treatment with modest doses of indomethacin. Extracts of the neoplasm, both primary and metastatic, contained large amounts of prostaglandins—both E and F. Prostaglandins were not detectable in blood, but this was attributed to lack of assay sensitivity.

Robertson et al (41) also described a patient with renal-cell carcinoma and hypercalcemia who had low plasma parathyroid concentrations but elevated immunoreactive PGE levels. Hepatic metastases contained higher concentrations of prostaglandins than normal kidney or liver. Ito et al (42) described a third hypercalcemic patient with renal adenocarcinoma who had a fall in serum calcium with indomethacin treatment. However, no measurements of prostaglandins were made.

In a more extensive study, Seyberth et al (43) reported on 29 patients with solid tumors. Fourteen were hypercalcemic (none of these 14 had renal carcinomas); 15 had normal blood calciums. Twelve of the 14 hypercalcemic patients had marked increases in urinary PG metabolites, while seven of the normocalcemic patients with cancer had elevations, and these were only slight to moderate. Six of the hypercalcemic patients were treated with indomethacin or aspirin. In all, urinary PG excretion fell, and there was a variable but significant decrease in serum calcium. Aspirin is a less potent inhibitor of PG synthesis than is indomethacin.

While the data just reviewed suggest that PG elaboration may be an extremely common cause of hypercalcemia associated with solid tumors, Tashjian (44) reported that a number of investigators have not found that a high percentage of patients respond to indomethacin. Thus, at the present level of understanding, solid carcinomas appear to produce hypercalcemia by elaboration of either parathormone or a PG-like material. The relative frequency of each remains to be elucidated.

A third mechanism for production of hypercalcemia is elaboration of osteoclast-stimulating factor (OSF). This material is a potent stimulator of osteoclast resorption of fetal bone in organ culture, found initially in supernatants of human peripheral leukocytes cultured with antigens or phytomitogens (45).

This material has been partially purified by Luben et al (46) and characterized as a peptide of molecular weight about 20,000. Weight for weight, OSF was as potent as purified bovine parathyroid hormone when tested for stimulation of bone resorption on fetal rat radius and ulna. Mundy et al (47) noted that hypercalcemia is common in patients with multiple myeloma, but rare in patients with diseases such as chronic lymphocytic leukemia, acute leukemia, or Waldenström's macroglobulinemia, in which the mass of intramedullary neoplastic cells is often as great as in myeloma. They assayed myeloma cells for OSF and found large quantities in the cells and in supernatants of cultures. Prostaglandins were not detectable in these same cells or supernatants. Parathormone was detectable by radioimmunoassay in myeloma cell extracts and cell culture supernatants. However, the amount of parathormone present was less than required to detectably increase bone resorption in the fetal bone bioassay. In contrast, the OSF content was adequate to explain the potent bone resorption properties of myeloma cell extracts or cell culture supernatants. These workers concluded that myeloma cells produce hypercalcemia by elaboration of OSF. Other white-cell malignancies associated with hypercalcemia may also elaborate OSF. Reported cases of lymphomas or leukemias associated with hypercalcemia are relatively rare, but have included acute lymphatic leukemia (48) and acute myeloblastic leukemia (49). Laird-Meyers (25) included 33 patients with lymphoma and 13 patients with leukemia (not further specified) among the 430 patients with hypercalcemia and malignancy he studied. In the patient with acute myeloblastic leukemia described by Zidar et al (50), in vitro cultures of the myeloid cells revealed presence or elaboration of parathormone similar to that described for the patient of Mundy et al (47). OSF was not quantified in the patient of Zidar, and they concluded that the myeloid cells produced hypercalcemia by elaboration of parathormone. In fact, serum parathormone was elevated. It is not possible to unravel these conflicting impressions at present, and conceivably some white cells may elaborate both peptides.

In summary, hypercalcemia associated with malignancies may be caused by elaboration of one or more of three humoral substances: parathormone, osteoclast-stimulating factor, and prostaglandins. The relative frequency of each as a cause is presently under study. In some cases elaboration of more than one of these substances may be causative.

GONADOTROPIN ELABORATION

Until some 10 years ago it was believed that extra pituitary sources of gonadotropin were relatively rare, confined to the syncytial and cytotrophoblastic cells of the pregnant primate, the endometrial cusps of the pregnant mare, neoplasms derived from the cytotrophoblast, teratomas containing such cells, and the hepatoblastoma (1, 51). In 1966 (52), 1967 (53), and 1968 (54), three case reports appeared describing patients with carcinoma of the lung elaborating gonadotropin. Subsequently, advances in knowledge of the peptide structure of the glycopeptide hormones LH, FSH, TSH, and HCG (15–20), have led to the elucidation that each of these four peptides is composed of two peptide chains, an alpha chain and a beta chain. As stated in the introduction, the alpha chain of each is very similar or identical in amino acid sequence; the beta chain is biochemically unique and confers biochemical, immunological, and biological specificity. Neither the alpha chain nor the beta chain has any biological potency (and neither shows, in vitro binding to gonadotropin receptors) (23). The beta chain of hCG and of LH are similar, differing in that hCG beta is longer by approximately 30 amino acids on the carboxyl end (19, 20, 55). The isolation and purification of the beta chains of these four peptides has permitted development of antisera directed against the beta chain, thus increasing specificity of radioimmunoassays (21, 22).

As indicated earlier, even using radioimmunoassays, which were much more sensitive than previous bioassays but could not distinguish LH from hCG, striking patients were occasionally reported in whom LH-hCG production quantitatively exceeded that possible by the pituitary, and a positive diagnosis of a neoplasm could be made based solely on this tumor marker (51–54). In two patients serum LH–hCG was elevated months prior to the ability to make a diagnosis of a neoplasm by conventional techniques (51, 54). When the more specific beta chain radioimmunoassay for hCG developed by Vaitukaitis et al (22) was applied to blood from patients with a variety of carcinomas, it was found that between six and 13 percent of the carcinomas were associated with increased blood concentration of hCG (56). Braunstein et al (56) reported that of 828 patients with nontesticular neoplasms studied, 60 (7%) had detectable hCG. Of particular interest were neoplasms of the gastrointestinal tract, which gave a higher frequency of positive values. The GI neoplasms studied, the number of patients, and the percent with detectable hCG in blood were as follows: gastric carcinoma, $N = 34$, 8 positive for hCG; colon-rectal, $N = 10$, 2 positive; pancreatic carcinoma, $N = 10$, 5 positive; hepatoma/hepatoblastoma, $N = 84$, 14 positive.

Subsequently, the story of hCG has evolved somewhat surprisingly. In 1960 Albert & Derner (57) reported that a large pool of menopausal urine contained a substance with biological properties similar to hCG. They based this on a detailed "biological fingerprint" of the urinary reference preparation (2nd IRPHMG), and contamination with pregnancy urine could not be excluded. In 1975 Braunstein (58) reported the presence of an hCG-like material in extracts of normal testes. To test the hypothesis that ectopic peptide elaboration is a universal concomitant of neoplasia (1), Odell and Wolfsen, et al, studied extracts of 87 carcinomas (saline and acetic acid extracts). All such extracts contained hCG as revealed by beta-hCG radioimmunoassay and by gonadotropin radioreceptor assay (3, 59). As control studies, extracts of liver and colon from patients without cancers were studied. Analysis by radioimmunoassay and radioreceptor assay showed that all contained a material indistinguishable from placental hCG (60, 61). The free beta chain of hCG does not react in the radioreceptor assay (but does react in the radioimmunoassay), and LH gives a dose-response line that is not parallel to hCG in the radioimmunoassay (but does give a parallel dose-response line to hCG in the radioreceptor assay). Thus the material in all carcinomas and in normal liver and colon studied could be identified as hCG, since it gave dose-response lines parallel to hCG in both radioimmunoassay and radioreceptor assay (3, 59–61). Chen et al (62) detected hCG-like material in urinary extracts of normal, nonpregnant human subjects.

Yoshimoto et al (60, 61) characterized the hCG present in normal tissues by concanavalin column studies. In such columns, peptides (such as placental hCG) containing sialic acid, galactose, and other carbohydrates bind to the column and are elutable by subsequent addition of methyl-α-glucopyranoside (MAGP) or borate. Carbohydrate-free peptides,which show little or no binding, elute in the void volume prior to application of MAGP. The hCG activity present in all extracts of normal colon and liver was shown to contain little or no carbohydrate (>90% eluted in the void volume) (60, 61). In contrast, over 90% of placental hCG showed binding to concanavalin. When the hCG activity present in carcinomas was studied, 30–50% contained carbohydrate as tested by concanavalin binding, and 50–70% was indistinguishable from the carbohydrate-free hCG present in normal tissues (59). Seventy percent of extracts of carcinomas contained amounts of this hCG activity indistinguishable from that present in normal tissues; approximately 30% contained amounts of hCG activity significantly greater than that in normal tissues.

Van Hall et al (63) and Tsuruhara et al (64) had previously shown that desialated hCG was rapidly cleared from the circulation and thus possessed little or no biological potency in vivo, though it possessed full immunological potency and receptor-binding potency in vitro.

Taking all these data into consideration, it seems reasonable to conclude that all or most normal tissues and all carcinomas synthesize the peptide sequence of hCG. The placenta and a portion of carcinomas add carbohydrate to the peptide sequence. In these two instances, metabolic clearance of the hCG is sufficiently slow that, if the carbohydrated hCG reaches the blood, it may be detectable by beta-hCG radioimmunoassay. For the majority of carcinomas and for normal tissues hCG is either not elaborated into blood, or if elaborated, metabolic clearance is so rapid that blood concentrations are not detectable. The function, if any, of this carbohydrate-free hCG in normal cells remains unknown.

Edmonds et al (65, 66) have shown that free alpha chain is secreted by the normal human pituitary along with the intact pituitary hormones FSH, LH, and TSH, i.e. alpha chain is secreted in excess. Since intact hCG is synthesized by (or at least is present in extracts of) normal tissues and carcinomas and elaborated into blood by about 10% of carcinomas, it seems reasonable that alpha chains may be commonly elaborated by cancers. In 1974 Rosen & Weintraub

(67) reported a single patient with metastatic gastric carcinoid, extracts of which contained very large quantities of free alpha chain. Increased alpha chain concentrations were found in blood of three other patients with bronchogenic carcinoma and three of 29 with carcinoid. Odell et al (3) and Wolfsen & Odell (68) studied the concentrations of alpha chain in blood of 186 patients with a variety of carcinomas. Forty-one percent had elevated concentrations when compared to age-matched controls. In these patients with carcinoma, hCG concentrations were only rarely increased in blood (6% of patients). It thus appears that alpha-chain elaboration by cancers may occur in the absence of intact hCG elaboration.

Gonadotropin production or free alpha-chain production by cancers produces few or no symptoms. In men gynecomastia may result from intact hCG production. In women, no clinical symptoms have been described, though in premenopausal women with carcinoma, irregular menses are not uncommon. Treatment of the gonadotropin or alpha-chain elaboration per se is not indicated.

HYPONATREMIA AND CANCER

Another syndrome associated with a variety of carcinomas consists of hyponatremia, hypervolemia, urine of inappropriately high osmolality (greater than plasma), and high urinary sodium concentrations. This syndrome is identical to that produced by continued administration of arginine vasopressin (AVP) and was originally ascribed to AVP elaboration by the carcinoma by Schwartz et al in 1957 (69). The syndrome was further defined by Bartter & Schwartz in 1967 (70). The majority of patients with this syndrome suffer from carcinoma of the lung, but within the past four to five years patients with other types of neoplasms have been reported. Squamous, bronchogenic, anaplastic, and oat-cell carcinomas of the lung have all been associated with this syndrome, as have carcinoma of the prostate (71), of the adrenal cortex (72), and Hodgkin's disease (73). Gilby (74) studied 49 patients with oat-cell carcinoma of the lung and reported in an abstract that 35% had syndromes of inappropriate secretion of vasopressin. In a full length manuscript Gilby et al (75) reported 40% of the series of oat-cell carcinomas studied were associated with sustained inappropriate vasopressin (SIADH) secretion.

The material elaborated by neoplasms associated with hyponatremia has been characterized by both bioassays and radioimmunoassays, and probably is AVP. Amatruda et al (76) and Bower & Mason (77) first demonstrated the nature of this material using bioassays. Subsequently, Vorherr et al (78) studied a series of patients; the vasopressin-like material in extracts of cancer reacted identically to AVP in a radioimmunoassay for vasopressin. Hirata et al (79) studied extracts of two carcinomas of the lung from patients with this syndrome. Using a bioassay, an immunoassay, and gel chromatography, the material was found to be identical or very similar to AVP. Recently Odell et al (3) characterized the vasopressin content of blood from 41 patients with carcinoma of the lung and 28 patients with carcinoma of the colon without clinical evidence of the SIADH syndrome. Forty-one percent of the patients with carcinoma of the lung and 43 percent with carcinoma of the colon had radioimmunoassayable vasopressin concentrations that were above those in a normal population dehydrated 12 hours. Presumably these patients would manifest the SIADH syndrome if given excess water.

In summary, approximately 40% of patients with carcinoma of the lung and a smaller percent of patients with a variety of other carcinomas have ectopic production of vasopressin.

Such patients may present to the physician with severe hyponatremia, psychosis, or seizures. Treatment is best given by severely restricting access to water. SIADH secretion produces no clinical symptoms unless excess free water is ingested. Other forms of treatment could include drugs that prevent the renal response to ADH, such as lithium carbonate or demeclocycline (demethylchlortetracycline).

HYPOGLYCEMIA AND CANCER

A wide variety of neoplasms may be associated with hypoglycemia. Symptoms of hypoglycemia may occur months or even years prior to recognition of the neoplasm, simultaneously with discovery, or after a well-established neoplasm is known. The types of neoplasms associated with hypoglycemia differ from those causing most other humoral cancer syndromes and are listed in Table 1. The largest number may be loosely defined as mesotheliomas (e.g. mesothelioma, fibrosarcoma, neuofibroma, sarcomatous dysembryoplasia, neurofibrosarcoma, mesenchymoma). These neoplasms are usually large at the time hypoglycemia is noted; size ranges from 800 to 10,000 g and averages about 2400 g. About two thirds of these mesotheliomas have occurred in the abdomen, both peritoneal and retroperitoneal. The remaining one third occur within the thorax. Fifty mesenchymal tumors producing hypoglycemia were distributed as follows: 24% retroperitoneally, 40% in the peritoneal area including pelvis, and 36% within the thorax. Sex incidence was equal (1, 80).

Table 1[a] The percentage of types of neoplasms producing hypoglycemia

Neoplasm	Percentage
Mesenchymal[b]	64
Hepatic	21
Adrenal carcinomas	6
Miscellaneous[c]	9

[a] Ectopic hormone syndromes only. This excludes insulinomas, nosidioblastosis, etc.

[b] Mesothelioma, fibrosarcoma, neurofibroma, neurofibrosarcoma, reticulosarcoma, lymphosarcoma, etc.

[c] Anaplastic carcinoma, adenocarcinomas of stomach and colon, pseudomyxoma, hypernephroma, lymphoma, hemangiopericytoma.

The hypoglycemia associated with hepatic carcinomas does not appear to be caused simply by direct tumor replacement of normal liver cells and resultant decrease in glucose production. Frequently, hypoglycemia is the presenting symptom prior to diagnosis of a neoplasm. Liver function in such a patient is often normal or nearly normal. Additionally, hypoglycemia is relatively common with hepatomas, but distinctly rare with severe hepatic infiltration from metastatic disease. As will be discussed subsequently, it is possible that by strict definition hypoglycemia caused by hepatoma is not an example of ectopic hormone production, but rather an example of a neoplasm retaining sufficient differentiation to produce a substance normally produced by hepatic cells.

Several cases of hypoglycemia associated with malignancies of white blood cells have been reported (81–83). In some cases this is artifactual, produced by collection of blood in tubes that permit continued glucose metabolism by the many circulating malignant cells (83). In other cases a true ectopic humoral syndrome appears to exist.

The etiology of the hypoglycemia remains poorly understood. The most likely explanation in most cases is tumor elaboration of a somatomedin. The somatomedins comprise a family of peptide hormones normally produced by the liver. Hepatic production of somatomedins is believed to be mediated via growth hormone, i.e. growth hormone stimulates hepatic production of one or more somatomedins. (It is in this sense that hepatic carcinomas that produce hypoglycemia may not represent an ectopic hormonal syndrome.) Somatomedins have biological actions identical to insulin and in fact have been purified using insulin receptor assays (84). Megyesi et al (85) reported elevated somatomedin (nonsuppressible insulin-like activity) by radioreceptor assay in five of seven patients with cancer and hypoglycemia.

In addition to the finding of Megyesi et al (85), other data suggest circumstantially that the material elaborated by neoplasms producing hypoglycemia is a somatomedin. For example, extracts from 25 neoplasms have been studied in an attempt to demonstrate a hypoglycemic-producing factor (reviewed in 1 and 80). Using bioassays (rat diaphragm or epididymal fat) in studies of 10 patients, insulin-like activity was demonstrated. Acid-alcohol extracts (used for insulin) generally contained significantly less bioactivity than saline, water, or buffer extracts, which suggests the material is not insulin per se. Radioimmunoassays for insulin of six extracts failed to reveal insulin activity; three of these same extracts had insulin-like activity revealed by bioassay (86–87). Based upon present knowledge, these substances with identical bioassay properties of insulin, and showing full reaction in insulin radioreceptor assays but no reaction in insulin radioimmunoassays, are somatomedins. However, further studies are required in additional patients to firmly establish somatomedins as the cause of hypoglycemia in nonbeta-cell cancers.

To further complicate the postulate that cancer elaboration of a somatomedin is the cause of hypoglycemia, two patients have been reported with elevated blood radioimmunoassayable insulin (88, 89). Kiang et al (88) reported a 25-year old woman with carcinoma of the cervix. Immunoassayable insulin was elevated in blood and in the primary and metastatic lesions. Honicky & de Papp (89) reported a five-year old hypoglycemic boy with a teratoma of the mediastinum, which contained beta cells. Preoperatively, plasma insulin levels were elevated (45 μU/ml), and postoperatively they were normal (15 μU/ml). These appear to be unusual patients, and other than these cases, insulin per se has not been reported to be elevated in a large number of patients studied.

ECTOPIC ACTH-MSH SYNDROME

The observation of Cushing's syndrome with adrenal hyperplasia in association with oat-cell carcinoma of the lung was made by Brown in 1928 (90). Following this, and particularly in the past few years, considerable advances in understanding the pathophysiology of the ectopic adrenocorticotropic hormone (ACTH) syndrome have been made.

Pathophysiology

In 1962, Liddle's group (91, 92) demonstrated that either primary or metastatic tumor tissue from patients with clinically recognizable ectopic ACTH syndrome contained high concentrations of immunoreactive ACTH. In patients harboring tumors not associated with clinically recognizable ectopic ACTH syndrome, ACTH that reacted in an immunoassay but possessed little bioactivity was described by Ratcliffe et al (93). Further characterization of this ACTH by Yalow & Berson (94, 95) and Gewirtz et al (96) showed that it was a larger form of ACTH with a molecular weight of about 20,000 and it was more acidic than ACTH_{1-39}. This "Big ACTH" has little bioactivity in the dispersed adrenal-cell bioassay, but brief treatment with trypsin releases bioactive "little ACTH" (96). These results have suggested that Big ACTH is a prohormone form of ACTH. Variable amounts of Big ACTH bind to concanavalin A–agarose columns and are eluted with carbohydrate competitors. The use of concanavalin was described also in the section on gonadotropin elaboration and indicates Big ACTH is a glycopeptide (97).

Parallel work on the biosynthesis of ACTH from mouse pituitary tumor-cell lines in culture and from purified messenger RNA from this source indicated that the first mRNA product with ACTH immunoreactivity has a molecular weight of 31,000. This is converted in minutes to a 23,000 dalton form, to a 13,000 dalton form, and finally to the 4,500 dalton form believed to be bioactive ACTH_{1-39} (98). In the whole cell system all but the 4,500 molecular weight form are glycosylated (7, 99). Thus the Big ACTH, constituting the major form of immunoreactive ACTH in tumors and a minor form of ACTH in human pituitaries, appears to be a normal biosynthetic precursor of ACTH_{1-39}. This proACTH has been found in significant concentration in virtually all extracts of lung carcinomas, regardless of histologic type (100), and in a variety of other tumor types including colon, breast, gastric, esophageal, and pancreatic (3). These data suggest that activation of the genome for ACTH precursor synthesis occurs commonly in carcinomas. The expression of the clinically apparent ectopic ACTH syndrome appears to be related to enzymatic ability of the tumor to convert proACTH to bioactive ACTH. In the peripheral blood of untreated lung cancer patients without clinical ectopic ACTH syndrome, proACTH is frequently elevated while $ACTH_{1-39}$ circulates in normal concentrations (6, 100, 101). ProACTH (bioinactive) is also the major component of ACTH immunoreactivity in the blood of patients with clinically apparent ectopic ACTH syndrome, but it does not circulate as a glycopeptide as far as is known (97). Bioactive ACTH also circulates in elevated concentrations in these patients. Increased production of bioactive ACTH results in adrenocortical hyperplasia, and glucocorticoid and androgen excess, producing Cushing's syndrome.

Tumors associated with the ectopic ACTH syndrome have also been shown by immunoassay and bioassay to produce melanocyte-stimulating hormone (MSH) (102–104). Human βMSH_{1-22} has recently been shown to be an artifact of extraction methods used for purification. This extraction method splits βMSH from larger peptides containing the entire βMSH sequence, the lipotropin (LPH) (104–108). βMSH per se does not exist in unextracted human plasma, or in heat-inactivated extracts of human pituitaries or tumors. βMSH activity in the form of the larger peptide, probably βLPH, has been demonstrated in more than 90% of tumor extracts from a variety of carcinoma types and is elevated in the blood of 61% of patients with untreated lung cancer without the ectopic ACTH syndrome (3). It is not surprising that plasma and tissue concentrations of ACTH and βMSH-βLPH closely correlate, since Phifer et al (109) observed that in the human these two peptides were synthesized by the same pituitary cell, and Roberts et al (110) reported that messenger RNA for ACTH directed synthesis of one large precursor molecule containing both ACTH and βLPH. Increased production of these two peptides together is believed to contribute to the hyperpigmentation observed in the ectopic ACTH syndrome.

The Clinical Syndrome

The secretion of large amounts of biologically active ACTH is clearly most commonly associated with carcinoma of the lung, particularly oat-cell type (111). Lung cancer accounts for about 50% of clinical cases (1, 2); bronchial carcinoids for about 5% of cases; epithelial tumors of the thymus for 10%; and pancreatic islet-cell, carcinoid, and adenocarcinoma for another 10%. Neoplasms from neural crest tissue (including pheochromocytoma, neuroblastoma, paraganglioma, and ganglioma) are the etiology in 5%, and medullary carcinoma of the thyroid is the cause in 5% of cases. However, any carcinoma should be considered a possible cause.

Because half of the patients with the ectopic ACTH syndrome have carcinoma of the lung with a rapidly fatal course, typical signs of Cushing's syndrome may not develop. Very high plasma cortisol levels produce weight loss, hypokalemic alkalosis, muscle wasting, edema, hypertension, impaired glucose tolerance, mental changes, and increased pigmentation.

Patients with slower-growing tumors such as thymomas, bronchial carcinoids, and pheochromocytomas develop the typical physical findings of Cushing's syndrome, in some cases years before the tumor is manifested (112). In patients with cancer elaboration of bioactive ACTH, hypokalemia is an important sign that suggests an ectopic ACTH source.

Diagnosis

In the clinical ectopic ACTH syndrome urinary free cortisol, 17 hydroxysteroics, and 17 ketosteroids show marked increase. Such increases in adrenal androgen production observed by very high dehydroepiandrosterone sulfate (DHEA-S) in plasma (113) and 17 ketosteroids in urine are seen only in adrenal carcinomas and ectopic ACTH syndrome. Plasma cortisol is frequently greater than 40μg%, plasma ACTH is greater than 200 pg/ml, and cortisol and ACTH typically do not suppress with 2 mg dexamethasone given every six hours for two days (114). Patients with pituitary-dependent Cushing's disease show over 50% suppression of plasma and urinary corticosteroids with this dose of dexamethasone. Patients with adrenal tumors have low baseline plasma ACTH. In about half of patients with Cushing's syndrome due to bronchial carcinoids (115), and in some with thymomas, reproducible suppression of cortisol or its metabolites with dexamethasone occurs. This phenomenon may be due to tumor production of corticotropin-releasing hormone (CRH) (116, 117), a small peptide normally produced by the hypothalamus to control pituitary ACTH secretion. In these patients with cancer who show suppression of blood cortisol, the dexamethasone probably acts at a pituitary level to inhibit CRH action. Alternatively, the dexamethasone could act directly at a tumor level to inhibit CRH production by binding with glucocorticoid receptors in the cytosol of tumor cells themselves (118), thus allowing suppressive effects of dexamethasone at the tumor-cell nucleus.

Periodic secretion of ACTH with a cycle of 10–18 days was reported (119, 120) and may present a difficult diagnostic problem.

Localization of occult tumors that produce Cushing's syndrome has been accomplished by selective venous sampling for ACTH; concentrations of ACTH are increased in venous drainage of the tumor (121). This diagnostic method also has been used to differentiate ectopic ACTH secretion from pituitary Cushing's syndrome (122).

The determination of significant amounts of proACTH in the blood may be useful in the diagnosis of ectopic ACTH production in the future.

Treatment

The treatment of choice for ectopic ACTH syndrome is complete surgical removal of the ACTH-secreting tumor. In nonresectable cases enzyme blockers (such as metyrapone), which inhibit cortisol, have been used acutely (123). The adrenocorticolytic drug o,p'-DDD requires several weeks before full effects are obtained; for this reason metyrapone may be used in the interim (124). These drugs may reduce blood cortisol to low levels, so replacement doses of glucocorticoids are required.

SYNDROMES OF GASTROINTESTINAL PEPTIDE PRODUCTION

There are several clinical syndromes caused by tumor production of gastrointestinal hormones. In the true sense these syndromes may not all be due to ectopic peptide production, they may represent peptide production by tumors derived from cells that normally synthesize these peptides.

The Zollinger-Ellison syndrome, as described in 1955 (125), consists of malignant peptic ulceration of the stomach, duodenum, and jejunum due to continuous maximal gastric-acid secretion stimulated by pancreatic-tumor secretion of gastrin (126). Severe diarrhea and even steatorrhea may occur (127). Tumors of the duodenal wall, antral-cell or islet-cell hyperplasia, and ovarian cystadenoma have been reported to produce this syndrome (128–132). Whether secretion of gastrin by pancreatic tumors represents ectopic gastrin production, or whether a gastrin-secreting cell normally exists in the pancreas, is debated (133, 134). These tumors have been found to contain big, big gastrin (amino acid sequence not yet characterized), big gastrin (34 amino acids), little gastrin (17 amino acids), minigastrin (13 amino acids), and to secrete these forms in response to secretin stimulation (135–137). Big gastrin, little gastrin and minigastrin are each bioactive and stimulate gastric acid secretion (137). However, neither the gastric acid hypersecretion nor the elevated serum gastrin explain the rapid rate of gastric emptying observed in this syndrome; other neural or humoral mechanisms appear to contribute to the clinical abnormalities (138). Immunohistochemical staining has shown more than one peptide-containing cell type in a significant percentage of tumors studied (139, 140), therefore multihormonal effects may be observed (141). The diagnosis of Zollinger-Ellison syndrome is made by finding elevated basal and calcium-stimulated serum-gastrin concentrations and a paradoxical rise in serum gastrin in response to secretin (142–146). Because the tumor is frequently multicentric or metastatic, the treatment of gastrinoma is total gastrectomy as well as removel of the primary and metastatic tumor, if possible (128, 147, 148). Total gastrectomy may be avoided if the patient is asymptomatic or if the tumor is well localized and gastric acid secretion halts during surgery following tumor resection (149). Prolonged survival and occasional tumor regression has been observed after total gastrectomy (150). Approximately 40% of patients in Zollinger's series have gastrinomas that behave as benign tumors (148). For nonresectable malignant gastrinomas, streptozotocin and 5-fluorouracil have been used to decrease tumor size and gastrin secretion (151).

Watery diarrhea with hypokalmeia and hypochlorhydria (Verner-Morrison syndrome) has been recognized as another humoral syndrome associated with pancreatic islet-cell carcinoma and other tumors (152–154). Tumor production of vasoactive intestinal polypeptide (155–157), prostaglandin E (158), human pancreatic polypeptide (140), secretin (159, 160), serotonin, enteroglucagon, pancreatic glucagon, and gastric-inhibitory polypeptide (161–163) have been reported with this symptom complex. Over half the patients with this syndrome have hypercalcemia, which returns to normal with removal of the tumor (164). If all of the tumor cannot be removed surgically, streptozotocin may be used as chemotherapy (165).

A clinical presentation of malabsorption with depressed small-intestinal motility and absorption was observed with a renal tumor that contained large amounts of enteroglucagon (166, 167). Enteroglucagon is immunologically distinct from pancreatic glucagon (168) and is normally produced by an intestinal mucosal cell that histologically resembles the pancreatic glucagon-secreting alpha cell (169).

There are two reported cases of hyperglycemia associated with pancreatic tumor secretion of somatostatin (170, 171). Somatostatin, a tetradecapeptide that suppresses glucagon and insulin secretion, is normally produced in the D cells of the pancreatic islets. Therefore, this also is not truly an instance of a tumor ectopically secreting somatostatin.

As has been implied repeatedly, preoperative diagnosis of benign vs malignant islet-cell tumors is often difficult or impossible. Kahn et al (172, 173) recently evaluated blood from 76 patients with functioning islet-cell tumors. The free alpha chain, the beta chain, or intact hCG (one or more of the peptides) was elevated in blood of 63% of patients with malignant tumors and in none of the patients with benign tumors. These gonadotropin tumor markers may prove highly useful in preoperative diagnosis of malignant vs benign islet-cell tumors.

HEMATOLOGIC ABNORMALITIES ASSOCIATED WITH TUMORS

Abnormalities of erythrocytes, white cells, and platelets have all been observed in association with, and possibly caused by, neoplasms. The best known of these is erythrocytosis, well recognized as a humoral paraneoplastic syndrome, which is reversed upon removal of the tumor. Of 340 cases reported (174), more than 50% were associated with renal carcinoma. Twenty percent were due to cerebellar hemangioblastomas, 15% to benign renal cysts and adenomas, 6% to uterine fibroids, and the remaining were associated with hepatomas, adrenal carcinomas, virilizing ovarian tumors, lung carcinomas, thymomas, and pheochromocytomas (174–176).

From 1–5% of patients with renal tumors and 9–20% of patients with cerebellar hemangioblastomas have polycythemia (175). Normally, the kidney is the major site of erythropoietin production (177). However, other tissues, particularly the liver (5, 178–180) are able to produce erythropoietin in response to severe hypoxia and anemia in anephric humans or animals. Erythropoietin production by renal or liver tumors may thus, strictly speaking, not represent ectopic hormone production.

There are several mechanisms by which tumors cause erythrocytosis. Hammond & Winnick (174) reviewed the reported cases of tumor-associated erythrocytosis; increased levels of erythropoietin (by bioassay) were observed in the blood of 24 of 25 cases prior to surgical removal of the tumor and in 1 of 13 postoperatively. Extracts of the tumor tissue or renal cystic fluid had increased erythropoietin activity compared to adjacent normal tissue in 28 of 44 tested. Erythropoietin activity in blood has been found to correlate with tumor activity and occurrence of metastases in individual patients (181).

The erythropoietin found in renal tumors or cysts, or in cerebellar cysts, is similar to erythropoietin normally present in anemic serum in electrophoretic mobility, molecular weight, and antisera neutralization studies (182). However, the elevated erythropoietin in the blood of patients with these tumors does not increase in response to phlebotomy, as it does in normal individuals (183).

A precursor or activator of erythropoietin has been reported in the light mitochrondrial and microsomal subcellular fractions of human kidneys; this precursor releases erythropoietin when incubated with normal serum (184, 185). Although not reported yet, it is possible that tumors elaborate an erythropoietin precursor, which may then become bioactive in blood.

A second mechanism by which tumors may produce erythrocytosis is exemplified by the adrenal cortical tumors. Serum and tumor extracts from three patients with erythrocytosis associated with these tumors did not have increased erythropoietin activity (186). These adrenal tumors and virilizing ovarian tumors (187) produce steroids that may have erythropoietic effects. Androgens have been observed to increase the activity of erythropoietin precursor or activator (184, 185). In addition, 5β-pregnane and 5β-androstane steroids appear to act directly on the non–erythropoietin responsive population of red-cell precursors to induce erythropoietin responsiveness (188).

A third mechanism for tumor-induced erythrocytosis is possible. Since prostaglandins have been shown to enhance the effects of erythropoietin on erythroid differentiation, it is possible that tumor elaboration of prostaglandins (already demonstrated to produce hypercalcemia) may result in erythrocytosis (189, 190). It is also possible in some cases that renal hypoxia is induced by the tumor. This would stimulate erythropoietin production by the normal renal cells.

Tumor associated erythrocytosis is diagnosed by abnormal elevations of red-cell mass without pancytosis or splenomegaly, as occurs in polycythemia rubra vera, and without decreased arterial oxygen saturation or hemoglobinopathy having increased hemoglobin-oxygen affinity. Demonstration of elevated erythropoietin activity in blood and urine aids in confirming the diagnosis. Surgical eradication of the tumor will result in disappearance of erythrocytosis. In inoperable cases, phlebotomy has been used to control the erythrocytosis.

Mild to marked elevations of granulocytes and monocytes have been observed in various types of neoplasms in the absence of infections (191–193). Robinson (194), using a granulopoietin purified from human urine as standard in a cell culture assay, reported elevated granulopoietic activity in the serum and urine of 12 patients with granulocytosis associated with a variety of tumors. However, no granulopoietic activity was found in extracts of two tumors studied. Therefore, there is little direct evidence for tumor production of this peptide. The granulopoietin produced by normal humans and isolated from urine is a glycoprotein with a molecular weight of about 40,000 (195).

Recently, an eosinophilopoietin has been described in mice; it is a small peptide with a molecular weight between 200 and 1300 (196). Although independent increase in eosinophils without increase in neutrophils has been observed in patients with neoplasmas, the relationship of the observation to the small peptide that stimulates eosinophil production in mice is not known. However, Parrillo et al (197) recently showed that patients with hypereosinophilic syndrome produce large amounts of eosinophilopoietin.

Platelets, fibrinogen, as well as factors V and VIII, have also been found to be elevated in patients with neoplasms (198, 199). Although intravascular coagulation and accelerated fibrinogen and platelet turnover have been observed (200–205), the relationship of thrombosis and bleeding to abnormal control of platelets is not yet clear. It is possible that tumor elaboration of several peptides related to the complement cascade or to other steps in coagulation control may be etiologic in these syndromes (2). However, they remain poorly understood, empirical, and clinical observations (204).

CANCER AND THE NEUROMUSCULAR SYSTEM

Most commonly, neurological symptoms produced by malignancies are caused by central nervous system metastases. Posner (206) estimated that approximately 16% of patients admitted to hospital with a malignancy have neurological symptoms. In addition to direct effects of cancer, however, a wide range of neurological disorders has been associated with carcinomas that are not metastatic to the nervous system. The cause of these is poorly understood, and for the most part they also remain at the level of empirical observations or correlations. Presumably in a very general sense they represent humoral syndromes caused by cancer. Brain and Norris (207) edited a splendid volume entitled *The Remote Effects of Cancer on the Nervous System* in 1964; Posner in 1971 (206), and Mawdsley in 1973 (208), reviewed these syndromes. While Croft & Wilkinson estimated that "humoral" neural and myopathic manifestations of carcinoma occur in 6.6% of patients with carcinoma, most of these patients have proximal muscle weakness, attributed by most neurologists to poor nutrition and weight loss, thus not representing a humoral syndrome. More uncommonly, central nervous or neuromuscular symptoms other than muscle weakness occur. These are usually associated with carcinoma of the lung, but also are not uncommonly seen in patients with carcinomas of the stomach, prostate, and ovary and malignancies of the myeloid and lymphatic cells; they are rarely associated with colon carcinoma. The neuromuscular disorders associated with malignancies include psychosis or dementia, demyelinating lesions, bulbar encephalitis, subacute cerebellar degeneration, optic neuritis, amyotrophic lateral sclerosis and long-tract degeneration in the spinal cord, various peripheral neuropathies, Guillain-Barre–type polyneuropathy, polymyositis, and myasthenic or myastheniagravis syndromes.

The ectopic ACTH syndrome may first manifest itself as psychosis. However, in the absence of excess bioactive ACTH production, patients (usually with lung carcinoma) may present with psychosis or dementia weeks or months before the neoplasm is discernable. These mental changes are not explainable by central nervous system metastases and presumably relate to some aspects of tumor metabolism or humor production.

Widely disseminated demyelinated lesions occurring with carcinomas have been commonly observed. Aström, Mancall & Richardson (209) called attention to this syndrome and suggested the term progressive multifocal leukoencephalopathy (PML) to describe it, even though the disease is actually more than a leukoencephalopathy, since gray matter as well as white matter is involved. Many of these patients with this syndrome suffer from chronic lymphocytic or myeloctyic leukemia, Hodgkin's disease, lymphosarcoma, or polycythemia vera. However, the lesions are also observed in association with carcinoma and with sarcoidosis and tuberculosis (207). The disease does not appear as an independent process or disease state. Of 44 patients reviewed by Richardson (207) 70% had some form of malignancy, and 20% had some form of non-neoplastic granulomatous disease. The disorder is characterized by the occurrence of widely disseminated demyelinating lesiona, i.e. disappearance of myelin sheaths with preservation of axis-cylinders. Typical lesions lie in the deep layers of the cortex. An additional striking finding in many (81%) of these patients is marked enlargement of the astrocytes. Clinically, the disorder generally progresses rapidly to death. The associated neoplasm may have been present several years prior to onset of PML.

One of the more common humoral neuromuscular disorders is subacute cerebellar degeneration (210). This remote effect of cancer may also precede the discovery of the malignancy by months to years. The clinical picture consists of bilateral and symmetrical cerebellar dysfunction with ataxia in arms and legs and dysarthria developing over a 3 to 6 week period. The cerebral spinal fluid is usually normal, but occasionally mild lymphocytosis and elevated protein exist (206). Pathology includes striking loss of Purkinje cells in a diffuse pattern throughout the cerebellum. In some patients degeneration of the long tracts of the spinal cord is also present. In others, associated degeneration of the nucleus of the IX, X, and XII nerves in the medulla and of the anterior horn cells of the spinal cord occurs. There is evidence that successful treatment of the malignancy reverses the disease, which generally stabilizes after 6–8 weeks and thereafter remains unchanged.

Ganglioradiculitis is manifested by subacute development of a distal sensory loss (especially proprioception) with relative preservation of motor power. Henson (211) described 11 patients with this disorder. Symptoms commonly precede the appearance of a malignancy, progress for several months, and then stabilize, leaving the patient with a severe disability.

Dermatomyositis or polymyositis is commonly associated with cancers (212, 213). Pearson (212) estimated that 20% of patients with this disease harbor a malignancy, usually a carcinoma and commonly intraabdominal. Shy & Silverstein (214) reported that 75% of men over 50 years old with polymyositis have a malignancy. This syndrome more directly relates to the neoplasm than subacute cerebellar degeneration or ganglioradiculitis, because successful treatment of the neoplasm frequently produces remission of the polymyositis.

Myasthenia gravis is commonly associated with thymoma. Lambert & Eaton (215) and Lambert & Rooke (216) described a different and most unusual "myasthenic" syndrome, which is usually associated with small-cell carcinomas of the lung. Clinically, the typical patient notes weakness and fatigue of proximal muscles, especially of the pelvic girdle and thighs, and in men, impotence. Muscle wasting is not present. In sharp contrast to myasthenia gravis and diagnostic of this syndrome, is the finding that strength increases over several seconds of sustained contraction, and on electromyography repeated nerve stimulation (at rates over 10 per second) causes a progressive increase in action potential. The symptoms do not respond to anticholinesterase medications, but often do respond with guanidine hydrochloride treatment (206). As with other humoral cancer syndromes, the presence of this syndrome mandates careful search for a neoplasm, in this instance a thoracic neoplasm.

ACKNOWLEDGMENTS

We are grateful to Ms. Elizabeth Stone for editing this manuscript.

Literature Cited

1. Odell, W. D. 1974. Humoral manifestations of non-endocrine neoplasms. In *Textbook of Endocrinology*, ed. R. H. Williams, pp. 1105–16. Philadelphia: Saunders. 1138 pp. 5th ed.
2. Odell, W. D., Wolfsen, A. 1975. Ectopic hormone secretion by tumors. In *Cancer: A Comprehensive Treatise*, ed. F. F. Becker, Vol. 3, pp. 81–98. New York: Plenum. 473 pp.
3. Odell, W. D., Wolfsen, A. F., Yoshimoto, Y., Weitzman, R., Fisher, D. A., Hirose, F. 1977. Ectopic peptide synthesis: A universal concomitant of neoplasia. *Trans. Assoc. Am. Phys.* 90:
4. Steiner, D. F., Oyer, P. E. 1967. The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. *Proc. Nat. Acad. USA* 57:473–80
5. Steiner, D. F. 1977. Insulin today. *Diabetes* 26:322–40
6. Gewirtz, G., Schneider, B., Krieger, D. T., Yalow, R. S. 1974. Big ACTH: Conversion to biologically active ACTH by trypsin. *J. Clin. Endocrinol. Metab.* 38:227–30
7. Eipper, B. A., Mains, R. E., Guenzi, D. 1976. High molecular weight forms of adrenocorticotropic hormone are glycoproteins. *J. Biol. Chem.* 251:4121–26
8. Habener, J. F., Kemper, B., Potts, J. T., Rich, A. 1972. Proparathyroid hormone: Biosynthesis by human parathyroid adenomas. *Science* 178:630–33
9. Habener, J. F., Kemper, B., Potts, J. T. Jr., Rich, A. 1975. Pre-proparathyroid hormone identified by cell-free translation of messenger RNA from hyperplastic parathyroid tissue. *J. Clin. Invest.* 56:1328–33
10. Gregory, R. A., Tracy, H. J. 1972. Isolation of two "big" gastrins from Zollinger-Ellison tumour tissue. *Lancet* 2:797–99
11. Walsh, J. H., Debas, H. T., Grossman, M. I. 1974. Pure human big gastrin immunochemical properties, disappearance half-time and acid-stimulating action in dogs. *J. Clin. Invest.* 54:477–85
12. Noe, B. D., Bauer, G. E., Stettes, M. W., Sutherland, D. E. R., Natarian, J. D. 1975. Glucagon biosynthesis in human pancreatic islets: Preliminary evidence for a biosynthetic intermediate. *Horm. Metab. Res.* 7:314–22
13. Noe, B. D., Bauer, G. E. 1975. Evidence for sequential metabolic cleavage of proglucagon to glucagon in glucagon biosynthesis. *Endocrinology* 97:868–77
14. Deftos, L. J., Roos, B. A., Bronzert, D., Parthermore, J. G. 1975. Immunochemical heterogeneity of calcitonin in plasma. *J. Clin. Endocrinol. Metab.* 40:409–12
15. Pierce, J. G., Liao, T. H., Howard, S. M., Shome, B., Cornell, J. S. 1971. Studies on the structure of

thyrotropin: Its relationship to luteinizing hormone. *Recent Prog. Horm. Res.* 27:165–212

16. Bellisario, R., Carlsen, R. B., Bahl, O. P. 1973. Human chorionic gonadotropin. Linear amino acid sequence of the α subunit. *J. Biol. Chem.* 248:6796–6809
17. Shome, B., Parlow, A. F. 1974. Human follicle stimulating hormone (hFSH): First proposal for the amino acid sequence of the α-subunit (hFSHα) and first demonstration of its identity with the α-subunit of human luteinizing hormone (hLHα). *J. Clin. Endocrinol. Metab.* 39:199–202

17a. Shome, B., Parlow, A. F. 1974. Human follicle stimulating hormone: First proposal for the amino acid sequence of the hormone-specific β subunit (hFSHβ). *J. Clin. Endocrinol. Metab.* 39:203–5

18. Ward, D. N., Reichert, L. E. Jr., Liu, W. -K., Nahm, H. S., Hsia, J., Lamkin, W. M., Jones, N. S. 1973. Chemical studies of luteinizing hormone from human and ovine pituitaries. *Recent Prog. Horm. Res.* 29:533–61
19. Papkoff, H., Sairam, M. R., Farmer, S. W., Li, H. L. 1973. Studies on the structure and function of interstitial cell-stimulating hormone. *Recent Prog. Horm. Res.* 29:563–90
20. Canfield, R. E., Morgan, F. J., Kammerman, S., Bell, J. J., Agosto, G. M. 1971. Studies of human chorionic gonadotropin. *Recent Prog. Horm. Res.* 27:121–64
21. Binoux, M., Pierce, J., Odell, W. D. 1974. Radioimmunological characterization of human thyrotropin and its subunits: Applications for the measurement of human TSH. *J. Clin. Endocrinol. Metab.* 38:674–82
22. Vaitukaitis, J. L., Braunstein, G. D., Ross, G. T. 1972. A radioimmunoassay which specifically measures human chorionic gonadotropin in the presence of human luteinizing hormone. *Am. J. Obstet. Gynecol.* 113:751–58
23. Ross, G. T., Vaitukaitis, J. L., Robbins, J. 1971. Preparation and properties of antisera made to α- and β-subunits of human chorionic gonadotropin. In *Structure-Activity Relationships of Protein and Polypeptide Hormones*, ed. M. Margoulies, F. C. Greenwood, p. 153. Amsterdam: Excerpta Med. 249 pp.
24. Warwick, O. H., Yendt, E. R., Olin, J. S. 1961. The clinical features of hypercalcemia associated with malignant disease. *Can. Med. Assoc. J.* 85:719–23
25. Laird-Meyers, W. P. 1960. Hypercalcemia in neoplastic disease. *Arch. Surg.* 80: 308–18
26. Bender, R. A., Hansen, H. 1974. Hypercalcemia in bronchogenic carcinoma. A prospective study of 200 patients. *Ann. Intern. Med.* 80:205–8
27. Albright, F. 1941. Case records of Massachusetts General Hospital #27461. *N. Engl. J. Med.* 225:789–94
28. Tashjian, A. H., Levine, L., Munson, P. L. 1964. Immunochemical identification of parathyroid hormone in non-parathyroid neoplasms associated with hypercalcemia. *J. Exp. Med.* 119:467–84
29. Berson, S. A., Yalow, R. S. 1966. Parathyroid hormone in plasma in adenomatous hyperparathyroidism, uremia, and bronchogenic carcinoma. *Science* 153:907–9
30. Sherwood, L. M., O'Riordan, J. L. H., Aurbach, G. D., Potts, J. T. 1967. Production of parathyroid hormone by nonparathyroid tumors. *J. Clin. Endocrinol. Metab.* 27:140–46
31. Buckle, R. M., McMillan, M., Mallinson, C. 1970. Ectopic secretion of parathyroid hormone by a renal adenocarcinoma in a patient with hypercalcemia. *Br. Med. J.* 4:724–26
32. Lipsett, M. B., Odell, W. D., Rosenberg, L. E., Waldmann, T. A. 1964. Humoral syndromes associated with nonendocrine tumors. *Ann. Intern. Med.* 61:733–56
33. Ratcliffe, J. G., Rees, L. H. 1974. Clinical manifestations of ectopic hormone production. *Br. J. Hosp. Med.* 11:685–92
34. Federman, D., Sherwood, L. M., Potts, J. 1971. Case records of the Massachusetts General Hospital. *N. Engl. J. Med.* 284:839–47
35. Powell, D., Singer, F. R., Murray, T. M., Minkin, C., Potts, J. T. 1973. Nonparathyroid humoral hypercalcemia in patients with neoplastic diseases. *N. Engl. J. Med.* 289:176–81

36. Tashjian, A. H., Voelkel, E. F., Levine, L. 1972. Evidence that the bone resorption–stimulating factor produced by mouse fibrosarcoma cells in prostaglandin E_2. A new model for the hypercalcemia of cancer. *J. Exp. Med.* 136:1329–43
37. Voelkel, E. F., Tashjian, A. H. Jr., Franklin, R., Wasserman, E., Levine, L. Hypercalcemia and tumor-prostaglandins: The VX_2 carcinoma model in the rabbit. *Metabolism* 24:973–86
38. Klein, D. C., Raisz, L. G. 1970. Prostaglandins: Stimulation of bone resorption in tissue culture. *Endocrinology* 86:1436–40
39. Robinson, D. R., Tashjian, A. H., Levine, L. 1975. Prostaglandin-stimulated bone resorption by rheumatoid synovia. A possible mechanism for bone destruction in rheumatoid arthritis. *J. Clin. Invest.* 56:1181–88
40. Brereton, H. D., Halushka, P. V., Alexander, R. W., Mason, D. M., Keiser, H. R., DeVita, V. T. Jr. 1975. Indomethacin-responsive hypercalcemia in a patient with renal-cell adenocarcinoma. *N. Engl. J. Med.* 291:83–85
41. Robertson, R. P., Baylink, D. J., Marini, J. J., Adkison, H. W. 1975. Elevated prostaglandins and suppressed parathyroid hormone associated with hypercalcemia and renal cell carcinoma. *J. Clin. Endocrinol. Metab.* 41:164–67
42. Ito, H., Sanada, T., Katayama, T., Shimazaki, J. 1975. Indomethacin-responsive hypercalcemia. *N. Engl. J. Med.* 293:558–59
43. Seyberth, H. W., Segre, G. V., Morgan, J. L., Sweetman, B. J., Potts, J. T., Oates, J. A. 1975. Prostaglandins as mediators of hypercalcemia associated with certain types of cancer. *N. Engl. J. Med.* 293:1278–83
44. Tashjian, A. H. Jr. 1975. Prostaglandins, hypercalcemia and cancer. *N. Engl. J. Med.* 293:1317–18
45. Horton, J. E., Raisz, L. G., Simmons, H. A., Oppenheim, J. J., Mergenhagen, S. E. 1972. Bone-resorbing activity in supernatant fluid from cultured human peripheral blood leukocytes. *Science* 177:793–95
46. Luben, R. A., Mundy, G. R., Trummel, C. L., Raisz, L. G. 1974. Partial purification of osteoclast-activating factor from phytohemagglutinin-stimulated human leukocytes. *J. Clin. Invest.* 53:1473–80
47. Mundy, G. R., Raisz, L. G., Cooper, R. A., Schechter, G. P., Salmon, S. E. 1974. Evidence for the secretion of an osteoclast-stimulating factor in myeloma. *N. Engl. J. Med.* 291:1041–46
48. Mawdsley, C., Holman, R. L. 1957. Hypercalcemia in acute leukaemia. *Lancet* 1:78–80
49. Kronfield, S. J., Reynolds, T. B. 1964. Leukemia and hypercalcemia: Report of a case and review of the literature. *N. Engl. J. Med.* 271:399–401
50. Zidar, B. L., Shadduck, R. K., Winkelstein, A., Zeigler, Z., Hawker, C. D. 1976. Acute myeloblastic leukemia and hypercalcemia. A case of a probable ectopic parathyroid hormone production. *N. Engl. J. Med.* 295:692–94
51. Rudnick, P., Odell, W. D. 1971. In search of a cancer. *N. Engl. J. Med.* 284:405–8
52. Fusco, F. D., Rosen, S. W. 1966. Gonadotropin-producing anaplastic large-cell carcinomas of the lung. *N. Engl. J. Med.* 275:507–15
53. Faiman, C., Colwell, J. A., Ryan, R. J., Hershman, J. M., Shields, T. W. 1967. Gonadotropin secretion from a bronchogenic carcinoma. Demonstration by radioimmunoassay. *Engl. J. Med.* 277:1395–99
54. Rosen, S. W., Becker, C. E., Schlaff, S., Easton, J., Gluck, M. C. 1968. Ectopic gonadotropin production before clinical recognition of bronchogenic carcinoma. *N. Engl. J. Med.* 297:640–41
55. Carlsen, R. B., Bahl, O. P., Swaninathan, N. 1973. Human chorionic gonadotropin. Linear amino acid sequence of the β subunit. *J. Biol. Chem.* 248:6810–27
56. Braunstein, G. D., Vaitukaitis, J. L., Carbone, P. P., Ross, G. T. 1973. Ectopic production of human chorionic gonadotrophin by neoplasms. *Ann. Intern. Med.* 78:39–45
57. Albert, A., Derner, I. 1960. Contamination of gonadotropin standards. *J. Clin. Endocrinol. Metab.* 20:1411–16
58. Braunstein, G. D., Rasor, J., Wade, M. E. 1975. Presence in normal human testes of a chorionic-go-

nadotropin-like substance distinct from human luteinizing hormone. *N. Engl. J. Med.* 293:1339–43

59. Yoshimoto, Y., Odell, W. D., Wolfsen, A. 1977. Glycosylation: The variable parameter in recognition of hCG production by cancers. *N. Engl. J. Med.* Submitted
60. Yoshimoto, Y., Wolfsen, A. R., Odell, W. D. 1977. *Clin. Res.* 25: 151 (Abstr.)
61. Yoshimoto, Y., Wolfsen, A. R., Odell, W. D. 1977. Human chorionic gonadotropin-like substance in nonendocrine tissues of normal subjects. *Science* 197:575–77
62. Chen, H. C., Hodgen, G. D., Matsuura, S., Lin, L. J., Gross, E., Reichert, L. E., Birken, S., Canfield, R. E., Ross, G. T. 1976. Evidence for a gonadotropin from nonpregnant subjects that has physical, immunological, and biological similarities to human chorionic gonadotropin. *Proc. Natl. Acad. Sci. USA* 73:2885–89
63. Van Hall, E. V., Vaitukaitis, J. L., Ross, G. T., Hickman, J. W., Ashwell, G. 1971. Immunological and biological activity of hCG following progressive desialation. *Endocrinology* 88:456–64
64. Tsuruhara, T., Dufau, M. L., Hickman, J., Catt, K. J. 1972. Biological properties of hCG after removal of terminal sialic acid and galactose residues. *Endocrinology* 91:296–301
65. Edmonds, M., Molitch, M., Pierce, J., Odell, W. D. 1975. Secretion of alpha subunits of luteinizing hormone by the anterior pituitary. *J. Clin. Endocrinol. Metab.* 41:551–55
66. Edmonds, M., Molitch, M., Pierce, J., Odell, W. D. 1975. Secretion of alpha and beta subunits of TSH by the anterior pituitary. *Clin. Endocrinol.* 4:525–30
67. Rosen, S. W., Weintraub, B. D. 1974. Ectopic production of the isolated alpha subunit of the glycoprotein hormones. *N. Engl. J. Med.* 290:1441–47
68. Wolfsen, A. R. Odell, W. D. 1977. Elaboration of free alpha chain of glycopeptide hormones by cancers. Submitted, *Ann. Intern. Med.*
69. Schwartz, W. D., Bennett, W., Curelop, S., Bartter, F. 1957. A syndrome of renal sodium loss and hyponatremia probably resulting from inappropriate secretion of antidiuretic hormone. *Am. J. Med.* 23:529–42
70. Bartter, F. C., Schwartz, W. B. 1967. The syndrome of inappropriate secretion of antidiuretic hormone. *Am. J. Med.* 42:790–806
71. Sellwood, R. A., Spencer, J., Azzopardi, J. G., Wapnick, S., Welbourn, R. B., Kulatilake, A. E. 1969. Inappropriate secretion of antidiuretic hormone by carcinoma of the prostate. *Br. J. Surg.* 56:933–35
72. Falchuk, K. R. 1973. Inappropriate antidiuretic hormone-like syndrome associated with an adrenocortical carcinoma. *Am. J. Med. Sci.* 266: 393–95
73. Cassileth, P. A., Trotman, B. W. 1973. Inappropriate antidiuretic hormone in Hodgkin's Disease. *Am. J. Med. Sci.* 265:233–35
74. Gilby, E. D. 1976. *Br. J. Dis. Chest* 70:282 (Abstr.)
75. Gilby, E. D., Rees, L. H., Bondy, P. K. 1976. Proc. Int. Symp. Biol. Charact. Hum. Tumours, 6th, Copenhagen, 1975. In *Advances in Tumour Prevention, Detection and Characterization,* ed. W. Davis, C. Maltoni, linguistic ed. P. R. J. Burch, Vol. 3, p. 132. New York: Am. Elsevier 424 pp.
76. Amatruda, T. T., Mulrow, P. J., Gallagher, J. C., Sawyer, W. H. 1963. Carcinoma of the lung with inappropriate antidiuresis. *N. Engl. J. Med.* 269:544–49
77. Bower, B. F., Mason, D. M. 1964. *Clin. Res.* 12:121 (Abstr.)
78. Vorherr, H., Massry, S. G., Utiger, R. D., Kleeman, C. R. 1968. Antidiuretic principle in malignant tumor extracts from patients with inappropriate ADH syndrome. *J. Clin. Endocrinol. Metab.* 28:162–68
79. Hirata, Y., Matsukura, S., Imura, H., Yakura, T., Ihjima, S., Nagase, C., Itoh, M. 1976. Two cases of multiple hormone producing small cell carcinoma of the lung. *Cancer* 38:2575–82
80. Odell, W. D. 1968. Humoral manifestations of non-endocrine neoplasms. In *Textbook of Endocrinology,* ed. R. H. Williams, pp. 1211–22. Philadelphia: Saunders. 1258 pp. 4th ed.
81. Collip, P. J. 1972. Hypoglycemia and leukemia. *Pediatrics* 49:788–89
82. Buffet, C., Bonnefond, A., Mignon,

M., Hakim, J. 1974. Hypoglycemia during lymphosarcoma and reticulosarcoma. *Nouv. Presse Med.* 3: 181–84

83. Solomon, J. 1974. Case report: Spurious hypoglycemia and hyperkalemia in myelomonocytic leukemia. *Am. J. Med. Sci.* 267:359–63
84. Van Wyk, J. J., Underwood, L. E., Hintz, R. L., Clemmons, O. R., Voina, S. J., Weaver, R. P. 1974. The somatomedins: A family of insulin-like hormones under growth hormone control. *Recent Prog. Horm. Res.* 30:259–318
85. Megyesi, K., Kahn, C. R., Roth, J., Gordon, P. 1974. Hypoglycemia in association with extrapancreatic tumors: Demonstration of elevated plasma NSILA-S by a new radioreceptor assay. *J. Clin. Endocrinol. Metab.* 38:931–34
86. Field, J. B., Keen, H., Johnson, P., Herring, B. 1963. Insulin-like activity of non-pancreatic tumors associated with hypoglycemia. *J. Clin. Endocrinol. Metab.* 23:1229–36
87. Tranquada, R. E., Bender, A. B., Beigelman, P. M. 1962. Hypoglycemia associated with carcinoma of the cecum and syndrome of testicular feminization. *N. Engl. J. Med.* 266:1302–6
88. Kiang, D. T., Bauer, G. E., Kennedy, B. J. 1973. Immunoassayable insulin in carcinoma of the cervix associated with hypoglycemia. *Cancer* 31:801–5
89. Honicky, R., de Papp, E. W. 1973. Mediastinal teratoma with endocrine function. *Am. J. Dis. Child.* 126: 650–53
90. Brown, W. H. 1928. A case of pluriglandular syndrome: Diabetes of bearded women. *Lancet* 2:1022–23
91. Meador, C. K., Liddle, G. W., Island, D. P., Nicholson, W. E., Lucas, C. P., Nucton, J. G., Luetscher, J. A. 1962. Cause of Cushing's Syndrome in patients with tumors arising from "nonendocrine" tissue. *J. Clin. Endocrinol. Metab.* 22:693–703
92. Orth, D. N., Island, D. P., Nicholson, W. E., Abe, K., Woodham, J. P. 1968. Proc. Symp. Oak Ridge Assoc. Univ. 1967. In *Radioisotopes In Medicine: In Vitro Studies*, ed. R. L. Hayes, F. R. Goswitz, B. E. P. Murphy, p. 251. Oak Ridge, Tennessee: At. Energ. Comm., Div. Tech. Inf. 753 pp.
93. Ratcliffe, J. G., Knight, R. A., Besser, G. M. 1972. Tumour and plasma ACTH concentrations in patients with and without the ectopic ACTH syndrome. *Clin. Endocrinol.* 1:27–44
94. Yalow, R. S., Berson, S. A. 1971. Size heterogeneity of immunoreactive human ACTH in plasma and in extracts of pituitary glands and ACTH producing thymoma. *Biochem. Biophys. Res. Commun.* 44:439–45
95. Yalow, R. S., Berson, S. A. 1973. Characterizatics of "Big ACTH" in human plasma and pituitary extracts. *J. Clin. Endocrinol. Metab.* 36:415–23
96. Gewirtz, G., Schneider, B., Krieger, D. T., Yalow, R. S. 1974. Big ACTH: Conversion to biologically active ACTH by trypsin. *J. Clin. Endocrinol. Metab.* 38:227–30
97. Orth, D. N., Nicholson, W. E. 1977. High molecular weight forms of human ACTH are glycoproteins. *J. Clin. Endocrinol. Metab.* 44:214–17
98. Mains, R. E., Eipper, B. A. 1976. Biosynthesis of adrenocorticotropic hormone in mouse pituitary tumor cells. *J. Biol. Chem.* 251:4115–20
99. Eipper, B. A., Mains, R. E. 1977. Peptide analysis of high molecular weight ACTH. *Endocr. Soc. Program, 59th Ann. Meet.* p. 58
100. Ayvazian, L. F., Schneider, B., Gewirtz, G., Yalow, R. S. 1975. Ectopic production of big ACTH in carcinoma of the lung. *Ann. Rev. Resp. Dis.* 111:279–87
101. Wolfsen, A., Odell, W. D. 1977. *Clin. Res.* 25:502 (Abstr.)
102. Shimizu, N., Ogata, E., Nicholson, W. E., Island, D. P., Ney, R. L., Liddle, G. W. 1965. Studies on the melanotropic activity of human plasma and tissues. *J. Clin. Endocrinol. Metab.* 25:984–90
103. Abe, K., Nicholson, W. E., Liddle, G. W., Island, D. P., Orth, D. N. 1967. Radioimmunoassay of βMSH in human plasma and tissues. *J. Clin. Invest.* 46:1609–15
104. Abe, K., Nicholson, W. E., Liddle, G. W., Orth, D. N., Island, D. P. 1969. Normal and abnormal regulation of βMSH in man. *J. Clin. Invest.* 48:1581–84
105. Scott, A. P., Lowry, P. J. 1974.

Adrenocorticotrophic and melanocyte-stimulating peptides in the human pituitary. *Biochem. J.* 139:593–96

106. Bloomfield, G. A., Scott, A. P. 1974. β-melanocyte stimulating hormone. *Proc. R. Soc. Med.* 67:748–50
107. Bloomfield, G. A., Scott, A. P., Lowry, P. J., Gilkes, J. J. H., Rees, L. H. 1974. A reappraisal of human βMSH. *Nature* 252:492–93
108. Bachelot, I., Wolfsen, A. R., Odell, W. D. 1977. Pituitary and plasma lipotropins: Demonstration of the artifactual nature of βMSH. *J. Clin. Endocrinol. Metab.* 44:939–46
109. Phifer, R. F., Orth, D. N., Spicer, S. S. 1974. Specific demonstration of the human hypophyseal adrenocortico-melanotropic (ACTH/MSH) cell. *J. Clin. Endocrinol. Metab.* 39:684–92
110. Roberts, J. L., Phillips, M. A., Herbert, E. 1977. Messenger RNA directed synthesis of high molecular weight form of ACTH. See Ref. 99, p. 57
111. Azzopardi, J. G., Williams, E. D. 1968. Pathology of nonendocrine tumors associated with Cushing's Syndrome. *Cancer* 22:274–86
112. Flint, L. D., Jacobs, E. C. 1974. Belated recognition of adrenocorticotropic hormone-producing tumors in post-adrenalectomized Cushing's Syndrome. *J. Urol.* 112:688–92
113. Yahmaji, R., Ibayashi, H. 1969. Plasma dehydroepiandrosterone sulfate in normal and pathological conditions. *J. Clin. Endocrinol. Metab.* 29:272–78
114. Liddle, G. W., Nicholson, W. E., Island, D. P., Orth, D. N., Abe, K. Lowder, S. C. 1969. Clinical and laboratory studies of ectopic humoral syndromes. *Recent Prog. Horm. Res.* 25:283–305
115. Mason, A. M. S., Ratcliffe, J. A., Buckly, R. M., Mason, A. S. 1972. ACTH secretion by bronchial carcinoid tumors. *Clin. Endocrinol.* 1:3–25
116. Upton, G. V., Amatruda, T. T. Jr. 1971. Evidence for the presence of tumor peptides with corticotropin-releasing-factor–like activity in the ectopic ACTH syndrome. *N. Engl. J. Med.* 285:419–24
117. Yamamoto, H., Hirata, Y., Matsukura, S., Imura, H., Nakamura, M., Tanaka, A. 1976. Studies on ectopic ACTH-producing tumors. IV. CRF-like activity in tumour tissue. *Acta Endocrinol. (Copenhagen)* 82:183–92
118. Garroway, N. W., Orth, D. N., Harrison, R. W. 1976. Binding of cytosol receptor-glucocorticoid complexes by isolated nuclei of glucocorticoid-responsive and nonresponsive cultured cells. *Endocrinology* 98:1092–1100
119. Bailey, R. E., 1971. Periodic hormonogenesis: A new phenomenon: Periodicity in function of a hormone producing tumor in man. *J. Clin. Endocrinol. Metab.* 32:317–27
120. Chajek, T., Romanoff, H. 1976. Cushing's Syndrome with cyclical edema and periodic secretion of corticosteroids. *Arch. Intern. Med.* 136:441–43
121. Rees, L. H. 1977. ACTH, lipotropin and MSH in health and disease. *Clin. Endocrinol. Metab.* 6:137–54
122. Corrigan, D. F., Schaaf, M., Whaley, R. A. Czerwinski, C. L., Earll, J. M. 1977. Selective venous sampling to differentiate ectopic ACTH secretion from pituitary Cushing's Syndrome. *N. Engl. J. Med.* 296:861–62
123. Orth, D. N., Liddle, G. W. 1971. Results of treatment in 108 patients with Cushing's Syndrome. *N. Engl. J. Med.* 285:243–47
124. Carey, R. M., Orth, D. N., Hartmann, W. H. 1973. Malignant melanoma with ectopic production of adrenocorticotropic hormone: Palliative treatment with inhibitors of adrenal steroid biosynthesis. *J. Clin. Endocrinol. Metab.* 36:482–87
125. Zollinger, R. M., Ellison, E. H. 1955. Primary peptic ulcerations of the jejunum associated with islet cell tumors of the pancreas. *Ann. Surg.* 142:709–23
126. McGuigan, J. E., Trudeau, W. L. 1968. Immunochemical measurement of elevated levels of gastrin in the serum of patients with pancreatic tumors of the Zollinger-Ellison variety. *N. Engl. J. Med.* 278:1308–13
127. Moshal, M. G., Brottman, S. A., Zameiter, N. 1969. *Clin. Res.* 17:308 (Abstr.)
128. Hofman, J. W., Fox, P. S., Wilson, S. D. 1973. Duodenal wall tumors and the Zollinger-Ellison syndrome.

surgical management. *Arch. Surg.* 107:334–39

129. Thompson, J. C., Hirose, F. M., Lemmi, C. A. E., Davidson, W. D. 1968. Zollinger-Ellison syndrome in a patient with multiple carcinoid islet cell tumors of the duodenum. *Am. J. Surg.* 115:177–84
130. Vance, J. E., Stoll, R. W., Kitabchi, A. E., Buchanan, K. D., Hollander, D., Williams, R. H. 1972. Familial nesidioblastosis as the predominant manifestation of multiple endocrine adenomatosis. *Am. J. Med.* 52:211–27
131. Friesen, S. R., Schimke, R. N., Pearse, A. G. E. 1972. Genetic aspects of the Zollinger-Ellison syndrome: Prospective studies in two kindred antral gastrin cell hyperplasia. *Ann. Surg.* 176:370–83
132. Cocco, A. E., Conway, S. J. 1975. Zollinger-Ellison syndrome associated with ovarian mucinous cystadenocarcinoma. *N. Engl. J. Med.* 293:485–86
133. Mitschke, H. 1977. Pathomorphologic studies of the endocrine cells in the gastrointestinal mucosa. Physiology, cytochemistry and ultra structure. *Veroeff. Morphol. Pathol.* 104:1–203
134. Lomsky, R., Langer, F., Vortel, V. 1969. Immunohistochemical demonstration of gastrin in mammalian islets of Langerhans. *Nature* 223:618–19
135. Dockray, G. J., Walsh, J. H., Passaro, E. Jr. 1975. Relative abundance of big and little gastrins in the tumors and blood of patients with the Zollinger-Ellison syndrome. *Gut* 16:353–58
136. Gregory, R. A. 1976. Heterogeneity of gastrins in blood and tissue. *Ciba Found. Symp.* 41:251–65
137. Walsh, J. H., Grossman, M. I. 1975. Gastrin. *N. Engl. J. Med.* 292:1324–34, 1377–84
138. Dubois, A., Eerdewegh, P. V., Gardner, J. D. 1977. Gastric emptying and secretion in Zollinger-Ellison syndrome. *J. Clin. Invest.* 59:255–63
139. Larsson, L. I., Grimelius, L., Akanson, R., Rehfeld, J. F., Stadel, F., Holst, J., Angervall, L., Sundler, F. 1975. Mixed endocrine pancreatic tumors producing several peptide hormones. *Am. J. Pathol.* 79:271–84
140. Larsson, L. I., Schwartz, T., Lundqvist, G., Chance, R. E., Sundler, F., Rehfeld, J. F. Grimelius, L., Fahrenkrug, J., Schaffalitzky de Muckadell, O., Moon, N. 1976. Occurrence of human pancreatic polypeptide in pancreatic endocrine tumors. Possible implication in the watery diarrhea syndrome. *Am. J. Pathol.* 85:675–84
141. Judge, D. M., Demers, L. M., Nahrwold, D. L., Dickman, P. S., Petrokubi, R. J., Trapukdi, S. 1977. Vasoactive intestinal polypeptide and gastrin producing islet cell carcinoma. *Arch. Pathol. Lab. Med.* 101:262–65
142. Friesen, S. R., Bolinger, R. W., Pearse, A. G. E., McGuigan, J. E. 1970. Serum gastrin levels in malignant Zollinger-Ellison syndrome after total gastrectomy and hypophysectomy. *Ann. Surg.* 172:504–21
143. Thompson, J. C. Reeder, D. D., Bunchman, H. H. 1972. Clinical role of serum gastrin measurements in the Zollinger-Ellison syndrome. *Am. J. Surg.* 124:250–61
144. Sanzenbacher, L. J., King, D. R., Zollinger, R. M. 1973. Prognostic implications of calcium mediated gastrin levels in the ulcerogenic syndrome. *Am. J. Surg.* 125:116–21
145. Morrow, D. J., Passaro, E. Jr. 1975. Calcium infusion test before and after total gastrectomy in the Zollinger-Ellison syndrome. *Am. J. Surg.* 129:62–66
146. Bradley, E. L. III, Galambos, J. T. 1976. Diagnosis of gastrinoma by the secretin suppression test. *Surg. Gynecol. Obstet.* 143:784–88
147. Zollinger, R. M., Martin, E. W. Jr., Carey, L. C., Sparks, J., Menton, J. P. 1976. Observations on the postoperative tumor growth behavior of certain islet cell tumors. *Ann. Surg.* 184:525–30
148. Fox, P. S., Hofmann, J. W., Wilson, S. D., DeCosse, J. J. 1974. Surgical management of the Zollinger-Ellison syndrome. *Surg. Clin. of North Am.* 54:395–407
149. Thompson, J. C., Reider, D. D., Villar, H. V., Fender, H. R. 1975. Natural history and experience with diagnosis and treatment of the Zollinger-Ellison syndrome. *Surg. Gynecol. Obstet.* 140:721–39
150. Friesen, S. R. 1967. Effect of gastrectomy on the Zollinger-Ellison tumor. Observations by second look procedures. *Surgery* 62:609–13

151. Cryer, P. E., Hill, G. J. II. 1976. Pancreatic islet cell carcinoma with hypercalcemia and hypergastrinemia: Response to streptozotocin. *Cancer* 38:2217–21
152. Said, S. I., Faloona, G. R. 1975. Elevated plasma and tissue levels of vasoactive intestinal polypeptide in the watery-diarrhea syndrome due to pancreatic, bronchogenic and other tumors. *N. Engl. J. Med.* 293:155–60
153. Verner, J. V., Morrison, A. B. 1958. Islet cell tumor and a syndrome of refractory watery diarrhea and hypokalemia. *Am. J. Med.* 25:374–80
154. Verner, J. V., Morrison, A. B. 1974. Endocrine pancreatic islet disease with diarrhea. *Arch. Intern. Med.* 133:492–500
155. Judge, D. M., Demers, L. M., Nahrwold, D. L., Dickman, P. S., Petrokubi, R. J., Trapukdi, S. 1977. *Arch. Pathol. Lab. Med.* 101:262–65
156. Gagel, R. F., Costanza, M. E., DeLellis, R. A., Norton, R. A., Bloom, S. R., Miller, H. H., Ucci, A., Nathanson, L. 1976. Streptozotocin-treated Verner-Morrison syndrome: Plasma vasoactive intestinal peptide and tumor responses. *Arch. Intern. Med.* 136: 1429–35
157. Said, S. I. 1975. Vasoactive intestinal polypeptide elevated plasma and tissue levels in the watery-diarrhea syndrome due to pancreatic and other tumors. *Trans. Assoc. Am. Physicians* 88:87–93
158. Jaffe, B. M., Condon, S. 1976. Prostaglandins E and F in endocrine diarrheagenic syndromes. *Ann. Surg.* 184:516–24
159. Cataland, S., Zollinger, R. M. 1970. *Clin. Res.* 18: 357 (Abstr.)
160. Bradley, E. L., Galambos, J. T. 1972. Secretin and Zollinger-Ellison syndrome. *Lancet* 1:594
161. Elias, E., Bloom, S. R., Welbourn, R. B., Kuzio, M., Polak, J. M., Pearse, A. G. E., Booth, C. C., Brown, J. C. 1972. Pancreatic cholera due to production of gastric inhibitory polypeptide. *Lancet* 2: 791–93
162. Burkhardt, A. 1976. The Verner-Morrison syndrome. The clinical picture and pathologic anatomy. *Klin. Wochenschr.* 54:1–11
163. Schmett, M. G. Jr., Soergel, K. H., Hensley, G. T., Chey, W. Y. 1975. Watery diarrhea associated with pancreatic islet cell carcinoma. *Gastroenterology* 69:602–16
164. Fisher, J. A., Blum, J. W., Binswanger, J. 1973. *Clin. Res.* 21:623 (Abstr.)
165. Kahn, C. R., Levy, A. G., Gardner, J. D., Miller, J. V., Gorden, P., Schein, P. S. 1975. Pancreatic cholera: Beneficial effects of treatment with streptozotocin. *N. Engl. J. Med.* 292:941–45
166. Gleeson, M. H., Bloom, S. R., Polak, J. M., Henry, K., Dowling, R. H. 1971. Endocrine tumor in kidney affecting small bowel structure mobility and absorptive function. *Gut* 12:773–82
167. Bloom, S. R. 1972. An enteroglucagon tumor. *Gut* 13:520–23
168. Unger, R. H., Ketterer, H., Eisentraut, A. M. 1966. Distribution of immunoassayable glucagon in gastrointestinal tissues. *Metabolism* 15:865–67
169. Polak, J. M., Bloom, S., Couling, I., Pearse, A. G. E. 1971. Immunofluorescent localization of enteroglucagon cells in the gastrointestinal tract of the dog. *Gut* 12:311–18
170. Ganda, O. T., Weir, G. G., Soeldner, J. S., Legg, M. A., Chick, W. L., Patel, Y. C., Ebeid, A. N., Gabbay, K. H., Reichlin, S. 1977. Somatostatinoma: A somatostatin-containing tumor of the endocrine pancreas. *N. Engl. J. Med.* 296:963–67
171. Larsson, L. I., Hirsch, M. A., Holst, J. J., Ingemansson, S., Kühl, C., Jensen, S. L., Lundqvist, G., Rehfeld, J. F., Schwartz, T. W. 1977. Pancreatic somatostatinoma. Clinical features and physiologic implications. *Lancet* 1:666–68
172. Kahn, C. R., Rosen, S. W., Weintraub, B. D., Fajans, S. S., Gorden, P. 1977. *Clin. Res.* 25:464 (Abstr.)
173. Kahn, C. R., Rosen, S. W., Weintraub, B. D., Fajans, S. S., Gorden, P. 1977. Ectopic production of chorionic gonadotropin and its subunits by islet cell tumors: A specific marker for malignancy. *N. Engl. J. Med.* 297:565–69
174. Hammond, D., Winnick, S. 1974. Paraneoplastic erythrocytosis and ectopic erythropoietins. *Ann. NY Acad. Sci.* 230:219–27
175. Valentine, W. N., Hennessy, T. G., Lang, E., Longmire, R., McMillan, R., Odell, W., Ross,

J. F., Scott, J. L., Simmons, D. H., Tanaka, K. R. 1968. Polycythemia: Erythrocytosis and erythremia. *Ann. Intern. Med.* 69:587–606

176. Balcerzak, S. P., Bromberg, P. A. 1975. Secondary polycythemia. *Semin. Hematol.* 12:353–82
177. Fisher, J. W., Birdwell, B. J. 1961. The production of an erythropoietic factor by the *in situ* perfused kidney. *Acta Haemat.* 26:224–32
178. Schooley, J. C., Mahlmann, L. J. 1974. Extra renal erythropoietin production by the liver in the weanling rat. *Proc. Soc. Exp. Biol. Med.* 145:1081–83
179. Peschle, C., Rappaport, I. A., Jori, G. P., Ghiariello, M., Gordon, A. S., Condorelli, M. 1973. Sustained erythropoietin production in nephrectomized rats subjected to severe hypoxia. *Br. J. Haematol.* 25:187–93
180. Naets, J. P., Wittek, M. 1968. Presence of erythropoietin in the plasma of one anephric patient. *Blood* 31:249–51
181. Murphy, G. P., Allen, J. E., Staubitz, W. J., Sinks, L. F., Mirand, E. A. 1972. Erythropoietin levels in patients with Wilms' tumor. *NY State J. Med.* 72:487–89
182. Waldmann, T. A., Rosse, W. F. 1964. Tumors producing erythopoiesis-stimulating factors. In *Hemoglobin, Its Precursors and Metabolism*, ed. F. W. Sundermann, F. W. Sundermann Jr., pp. 276–80. Philadelphia: Lippincott
183. Adamson, J. W. 1968. The erythropoietin/hematocrit relationship in normal and polythemic man: Implications of marrow regulation. *Blood* 32:597–608
184. Kuratowska, Z., Lewartowski, B., Lipinski, B. 1964. Chemical and biologic properties of an erythropoietin-generating substance obtained from perfusates of isolated anoxic kidneys. *J. Lab. Clin. Med.* 64:226–37
185. Gordon, A. S., Cooper, G. W., Zanjani, E. D. 1967. The kidney and erythropoiesis. *Semin. Hematol.* 4:337–58
186. Waldmann, T. A., Rosse, W. F., Swarm, R. L. 1968. The erythropoiesis-stimulating factors produced by tumors. *Ann. NY Acad. Sci.* 149:509–15
187. Gardner, F. H. 1962. Polycythemia. *Dis. Mon.* June:1–62
188. Singer, J. W., Samuels, A. I., Adamson, J. W. 1974. Steroids and erythropoiesis: *In vitro* evidence of different target cells for etiocholamolane and fluoxy mesterone. *Proc. Ann. Meet. Am. Soc. Hematol., 17th*, p. 46.
189. Dukes, P.P. 1971. *Blood* 38:822 (Abstr.)
190. Dukes, P. P. 1972. *Fed. Proc.* 31:487 (Abstr.)
191. Barrett, O. Jr. 1970. Monocytosis in malignant disease. *Ann. Intern. Med.* 73:991–92
192. Fahey, R. J. 1951. Unusual leukocyte response in primary carcinoma of the lung. *Cancer* 4:930–35
193. Hilts, S. V., Shaw, C. C. 1953. Leukemoid blood reaction (hyperleukocytosis) in malignancy. *Am. J. Clin. Pathol.* 12:218–22
194. Robinson, W. A. 1974. Granulocytosis in neoplasia. *Ann. NY Acad. Sci.* 230:212–18
195. Stanley, E. R., Metcalf, D. 1969. Partial purification and some properties of the factors in normal and leukaemic urine stimulating bone marrow colony growth *in vitro. Aust. J. Exp. Biol. Med. Sci.* 46:335–49
196. Mahmoud, A. A. F., Stone, M. K., Kellermeyer, R. W. 1977. *Clin. Res.* 25:519 (Abstr.)
197. Parrillo, J. E., Fauci, A. S., Wolff, S. M. 1977. *Clin. Res.* 25:519 (Abstr.)
198. Amundsen, M. A., Spittell, J. A. Jr., Thompson, J. H. Jr., Owen, C. A. Jr. 1963. Hypercoagulability associated with malignant disease and with the postoperative state: Evidence for elevated levels of antihemophilic globulin. *Ann. Intern. Med.* 58:608–16
199. Levin, J., Conley, C. L. 1964. Thrombocytosis associated with malignant disease. *Arch. Intern. Med.* 114:497–500
200. Joseph, R. R., Day, H. J., Sherwin, R. W., Schwartz, H. G. 1967. Microangiopathic anemia associated with consumption coagulopathy in a patient with disseminated carcinoma. *Scand. J. Haematol.* 4:271–82
201. Kenny, W. E. 1943. The association of carcinoma in the body and tail of the pancreas with multiple venous thrombi. *Surgery* 14:600–9
202. McKay, D. G., Wahle, G. H. 1955. Disseminated thrombosis in colon cancer. *Cancer* 8:970–78

203. McKay, D. G. 1965. *Disseminated Intravascular Coagulation: An Intermediary Mechanism of Disease*. New York: Hoeber Medical Division, Harper & Rowe. 493 pp.
204. Sproul, E. E. 1938. Carcinoma and venous thrombosis: Frequency of the body or tail of the pancreas with multiple venous thromboses. *Amer. J. Cancer* 34:566–85
205. Harker, L. A., Slichter, S. J. 1972. Platelet and fibrinogen consumption in man. *N. Engl. J. Med.* 287:999–1005
206. Posner, J. B. 1971. Neurological complications of systemic cancer. *Med. Clin. North Am.* 55:625–46
207. Croft, P. B., Wilkinson, M. 1965. The incidence of carcinomatous neuromyopathy with special reference to carcinoma of the lung and the breast. In *The Remote Effects of Cancer on the Nervous System*, ed. L. Brain, F. H. Norris, pp. 44–54. New York: Grune & Stratton. 230 pp.
208. Mawdsley, C. 1973. Neurological aspects of carcinoma and systemic diseases. *Practitioner* 211:282–94
209. Aström, K. E., Mancall, E. L., Richardson, E. P. 1958. Progressive multifocal leukencephalopathy; a hitherto unrecognized complication of chronic lymphatic leukemia and Hodgkin's disease. *Brain* 81:93–111
210. Brain, L., Wilkinson, M. 1965. Subacute cerebellar degeneration associated with neoplasms. *Brain* 88: 465–78
211. Henson, R. A., Hoffman, H. L., Urich, H. 1965. Encephalomyelitis with carcinoma. *Brain* 88:449–64
212. Pearson, C. M. 1966. Polymyositis. *Ann. Rev. Med.* 17:63–82
213. Bohan, A., Peter, J. B. 1975. Polymyositis and dermatomyositis. *N. Engl. J. Med.* 292:344–47, 403–7
214. Shy, C. M., Silverstein, I. 1965. A study of the effects upon the motor unit by remote malignancy. *Brain* 88:515–28
215. Eaton, L. M., Lambert, E. H. 1957. Electromyography and electric stimulation of nerves in diseases of the motor units. *J. Am. Med. Assoc.* 163:1117
216. Lambert, E. H., Rooke, E. D. 1965. Myasthenic state and lung cancer. See Ref. 207, pp. 67–80

Ann. Rev. Med. 1978. 29:407–15
Copyright © 1978 by Annual Reviews Inc. All rights reserved

MEDICAL TREATMENT OF ACROMEGALY

◆7288

Ira D. Goldfine, M.D.

Endocrinology-Metabolism Section, Medical Service, Veterans Administration
Hospital, and Department of Medicine, University of California, San Francisco,
California 94121

INTRODUCTION

Acromegaly, a disease resulting from a growth hormone(GH)–secreting adenoma of the pituitary gland, causes morbidity and mortality in affected patients by both local and peripheral physical changes (1–7). First, GH–secreting adenomas, like other tumors of the pituitary, cause compression of surrounding neural and bony structures, among which are the normal pituitary tissues, the sphenoid and cavernous sinuses, optic nerves, and third ventricle. Second, the excess secretion of GH in acromegaly can lead to changes in the peripheral tissues, which result in gross disfigurement of the face, hands, and other parts of the body; crippling arthritis; carpal tunnel syndromes; heart disease; hypertension; lung disease; a number of dermal changes; and carbohydrate and lipid abnormalities. Obviously, patients with acromegaly require therapeutic intervention to prevent or reverse both types of alterations.

The object of therapy for GH-secreting tumors and other pituitary tumors is to ablate the neoplasm and leave the normal pituitary tissues and surrounding structures intact. At present there are two readily available, definitive treatments for acromegaly: megavoltage x-ray radiation and selective transsphenoidal microsurgery[1] (8–13). Although both procedures can lead to marked reduction in circulating-GH levels in approximately 50–75% of patients so treated, neither form of therapy is ideal (8–13). Both forms of treatment may cause hypopituitarism and both have specific complications. For instance, radiation treatment has been associated with hair loss, temporal lobe epilepsy, and pituitary sarcomas (9).

In addition, the full results of radiation treatment may not become manifest for up to five years. In transsphenoidal surgery, the patient must undergo a major

---

[1] Heavy particle radiation is also a successful treatment modality for patients with acromegaly (4); this form of therapy, however, is available in only a few centers.

0066-4219/78/0401-0407$01.00

operative procedure that may result in cerebrospinal fluid rhinorrhea, meningitis, and even death (11–13). Thus an alternative form of treatment would be of great benefit. Medical (or pharmacologic) therapy, which is now becoming available, is a third option.

REGULATION OF GH SECRETION

Drug therapy of acromegaly is based on our current knowledge of the regulation of GH secretion. In normal man, GH is secreted by the α-cell of the pituitary (14). Stimulation of pituitary GH secretion is mediated by a factor termed GH-releasing factor (GRF), which is secreted into the hypothalamic-portal system by hypothalamic neurons (14–17). In turn, GRF release from the hypothalamus is believed to be regulated by at least three neurocenters: the ventromedial and arcuate nuclei of the hypothalamus and the limbic system of the brain (Figure 1; see also 14–17). Inhibition of pituitary GH secretion is mediated by the hypothalamic hormone, somatostatin (Figure 1; see also 18).

The Ventromedial Nucleus

The ventromedial nucleus does not stimulate GRF directly but appears to act through the arcuate nucleus (14). Agents that transmit signals from the ventro-

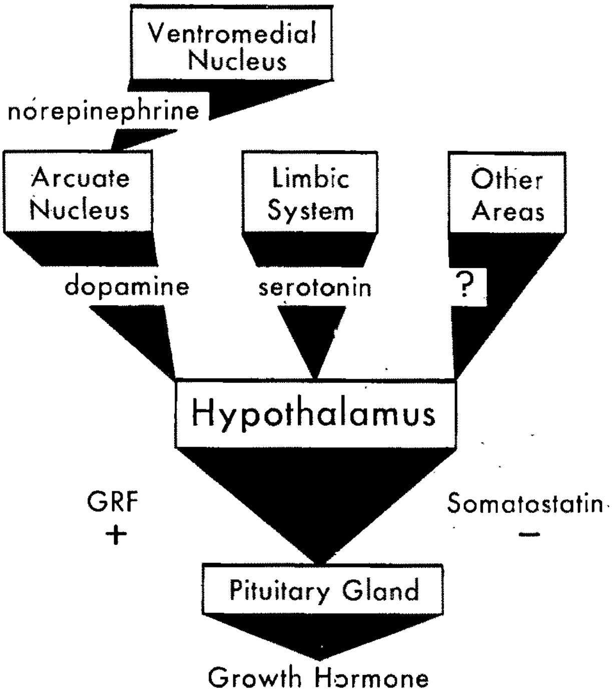


Figure 1 Simplified schematic diagram of the regulation of GH release (14–17).

medial nucleus to the arcuate nucleus of the hypothalamus are α-adrenergic catecholamines (Figure 1; see also 14–17). The ventromedial nucleus possesses a glucoreceptor that responds to falling glucose levels by stimulating GH release, and to rising glucose levels by decreasing GH release (14–17). In addition, stimuli such as glucagon, vasopressin, and arginine also appear to act via the ventromedial nucleus (14). Blockade of α-adrenergic receptors with agents such as phentolamine blunts the stimulatory effects of these substances (14).

The Arcuate Nucleus

The arcuate nucleus regulates GH release via the neurotransmitter dopamine (14–17). In normal man, L-dopa (a dopamine precursor), apomorphine, and 2-bromo-α-ergocryptine (dopamine agonists) stimulate GH release, whereas chlorpromazine, a postsynaptic inhibitor of the dopamine receptor, blocks GH release (14).

The Limbic System

The limbic system stimulates GH release via the neurotransmitter serotonin (14–17). This system plays an important role in slow-wave sleep, which is associated with a rise in GH levels. In addition, pyrogens may excite GH release via their actions on the limbic system (14).

Possible Regulation of GH Release via Other Neural Pathways

Stress, exercise, protein depletion, and estrogen administration are known to increase GH release, whereas other stimuli or metabolic states, such as obesity, fatty acids, medroxyprogesterone, and glucocorticoids, inhibit its release (14–17). Where these factors act is unknown. Thus, it is possible that areas in the nervous system, other than those mentioned above, also produce inputs into hypothalamic centers that in turn regulate the release of GRF.

REGULATION OF GH RELEASE IN ACROMEGALY

Ten years ago it was thought that GH production in acromegaly was the result of a pituitary adenoma that secreted GH autonomously, a concept that appeared to be supported by the early measurements of GH by radio immunoassay. Major changes in GH levels were not seen in the first patients studied, as GH levels generally were neither suppressed by glucose administration nor elevated by hypoglycemia (19–20). In 1969, however, Cryer & Daughaday (21) reported several patients with acromegaly whose GH levels fluctuated widely during the day and could be stimulated by either arginine infusion or hypoglycemia. These investigators suggested that, in some patients with acromegaly, GH release was not autonomous.

Subsequently, in an investigation of a larger number of acromegalics, Lawrence, Kirsteins, and I (22) found that, in almost every patient studied, some form of stimulus provoked a significant rise or fall in GH release. In many patients, we also noted paradoxical responses in GH levels; e.g. in certain patients, hypoglycemia produced a fall in GH levels, whereas hyperglycemia produced a rise. In any given acromegalic patient, the response to stimuli, such as arginine, hypoglycemia, or exercise, might be absent, normal, or paradoxical. Of major clinical importance has been the finding that L-dopa, which stimulates GH release in normal man, paradoxically inhibits GH release in acromegalics (23). The reasons for paradoxical responses are unknown, but they may represent abnormal receptor mechanisms in either the hypothalamus or pituitary.[2] From our observations, we postulated that in certain cases of acromegaly the secretion of GH by the pituitary tumor is not autonomous but rather can be regulated to some extent by the hypothalamus (22). We also suggested that certain cases of acromegaly may be caused by overstimulation of the pituitary by GRF.

Later studies by Hagen, Lawrence & Kirsteins (26) indicated that, in normal man, elevation of GH levels suppresses the stimulation of GH release by a variety of stimuli, such as hypoglycemia or arginine infusion. This suggested that GH normally inhibits its own release and this inhibition, therefore, is absent in patients with acromegaly. These investigators also detected the presence of excess levels of GRF in the serum of certain patients with acromegaly (27). Because these studies suggested that the hypothalamus may regulate GH release in certain patients with acromegaly, the possibility was raised that pharmacologic agents could decrease hypothalamic stimulation and thus ameliorate the excess GH secretion.

PHARMACOLOGIC THERAPY OF ACROMEGALY

Estrogens

It had been known for many years that estrogen administration decreases the signs and symptoms of acromegaly; but with the advent of the radio immunoassay of GH, it soon was clear that estrogens do not act by lowering serum GH levels (28). It was deduced that estrogen acts peripherally to antagonize the actions of GH on target tissues. Estrogen administration is helpful in certain female patients with acromegaly, but produces obvious adverse side effects in male patients.

Medroxyprogesterone

In 1967, Simon and co-workers (29) noted in normal subjects that a synthetic progesterone, medroxyprogesterone, blunted the rise of GH levels elicited by both arginine infusions and hypoglycemia. In 1971, Lawrence & Kirsteins (30) reported that, in a number of patients with acromegaly, the oral administration of medroxyprogesterone had a marked effect in both lowering GH levels and decreasing clinical symptoms. Subsequent studies confirmed that this agent was effective in certain patients with acromegaly, but its efficacy was not as broad as initially thought (31). Nonetheless, these studies provided the major conceptual impetus for all the subsequent studies of drug treatment of acromegaly. Medroxyprogesterone still remains a potential therapeutic agent in certain cases of acromegaly.

---

[2] For instance, both thyrotropin-releasing hormone (TRH) and gonadotropin-releasing factor have no effect on GH release in normal subjects but stimulate GH release in many acromegalics (24–25). This finding suggests that the GH-secreting cells in pituitary adenomas may have receptors for hypothalamic-releasing factors other than GRF.

Dopaminergic Drugs

The drug therapy of acromegaly is based on the observation that GH release from the pituitary is regulated by at least three hypothalamic centers. Two observations of the GH secretory patterns in acromegalics have led to a major advancement in the medical treatment of this disease. The first observation was that the GH secretion from tumors in many acromegalics was not autonomous; rather there seemed to be some degree of hypothalamic control. The second observation was that the response of GH secretion to stimuli could be paradoxical. Liuzzi and co-workers (23) first reported that the acute administration of L-dopa to acromegalics paradoxically produced a fall in GH levels, in contrast to normal individuals, in whom there was a rise. It now is known that a variety of dopaminergic drugs acutely lower GH levels in many acromegalic patients (32). These drugs include L-dopa and the ergot alkaloid, 2-bromo-α-ergocryptine (CB-154 Sandoz).

Since 2-bromo-α-ergocryptine has a prolonged effect in reducing GH levels in acromegaly, this drug has received the greatest number of clinical trials (33–38).[3] Studies from a number of investigators indicate that the long-term treatment of acromegalic patients with this drug leads to a significant fall in GH secretion in over half of the patients tested (33–38). In addition to lowered GH levels, in many patients there is improved glucose tolerance, a fall in urinary hydroxyproline excretion, and regression of the acroenlargement (33–38). The major side effects of this drug are nausea, vomiting, and hypotension, but these reactions usually are not serious problems when the drug is administered in gradually increasing doses. Other reported side effects are constipation, drug-induced alcohol intolerance, leg cramps, dry mouth, cold-induced digital spasm, and gastric ulcer. The starting dose of 2-bromo-α-ergocryptine usually is 2.5 mg per day, which is increased to 10–20 mg per day in divided dosages (38). Drug therapy must be continued in responsive patients; otherwise the elevated GH levels return. The acute testing of patients with either TRH or 2-bromo-α-ergocryptine does not necessarily predict the effects of chronic administration of this drug (38).

---

[3] It is of interest that this drug is used also to treat excess prolactin secretion from prolactin-secreting adenomas (39). In the latter instance, the effect is not paradoxical; dopamine probably is a natural prolactin-inhibiting factor, and in this case 2-bromo-α-ergocryptine is exerting a normal dopaminergic effect.

Chlorpromazine, an inhibitor of the dopamine receptor, was reported to lower GH levels in a single acromegalic patient (40). In a large series of patients, however, this drug was found to be ineffective (41).

α-Adrenergic Agents

As mentioned previously, α-adrenergic stimulation leads to release of GRF and subsequent stimulation of GH release from the normal pituitary. Phentolamine, an α-adrenergic antagonist, acutely lowers GH levels in acromegalic patients (42). The drug fenfluramine, a derivative of amphetamine, depletes cerebral amine concentrations, especially those amines located in the hypothalamus (43). Fenfluramine blunts the exercise-stimulated rise in GH release in normal subjects, and in a limited number of studies this drug also has been shown to lower GH levels in acromegalics (43). Drugs of this nature, in addition to dopaminergic agents, may be helpful in controlling GH release in acromegalics.

Antiserotonin Agents

Serotonin mediates the release of GRF via the limbic system. Several studies have indicated that serotonin antagonisis—especially metergoline and cyproheptadine, the latter, an effective agent in Cushing's disease—are also effective in lowering both basal and stimulated concentrations of GH in acromegalics (44–45).

Somatostatin

Somatostatin, a polypeptide hormone synthesized in the hypothalamus and other endocrine areas, markedly inhibits GH release as well as the release of other hormones such as insulin and glucagon (18). Unlike the other agents, somatostatin acts at the level of the pituitary. Injection or infusion of somatostatin to acromegalics leads to a profound fall in GH levels (46). At present there are several drawbacks to its use: it must be administered parenterally, it has a short half-life, it diminishes the secretion of other hormones such as glucagon and insulin, and it may be toxic. New analogs of somatostatin, however, are being developed that are relatively specific for GH-secreting cells. Further studies are necessary to determine the efficacy of these newer derivatives (47).

Other Agents

Alcohol, glucagon, and glucocorticoids inhibit GH secretion in patients with acromegaly, but their clinical potential remains uncertain (42, 48, 49).

CONCLUSION

Conventional x-ray radiation and/or transsphenoidal microsurgery still remain the major definitive treatments for acromegaly. It now appears, however, that certain patients can be treated successfully with pharmacologic agents that influence hypothalamic-pituitary relationships. One drug, 2-bromo-α-ergocryptine, can be of major benefit to many patients, but, unfortunately, it is not generally available in this country. It is reasonable to attempt the use of this drug or any of the other aforementioned agents in acromegalic patients who have failed the two definitive forms of therapy or are not candidates for surgery or radiation.

ACKNOWLEDGMENTS

I wish to acknowledge Dr. Ann M. Lawrence, whose continued dedication to the care and treatment of acromegalic patients has inspired many physicians to search for better ways of treating this disease. I thank Dr. T. J. Merimee for supplying me with material on the neural regulation of growth hormone release; Drs. R. Vigneri and M. D. Siperstein for their critical review of this manuscript; and Mrs. Vera Vaughn for editorial assistance.

This work was aided by the Medical Research Service of the Veterans Administration, the Northern California Affiliate, Inc. of the American Diabetes Association, by NIH grant AM-19415, and American Cancer Society grant BC-265.

The author regrets that space limitations preclude reference to every major study in the area of medical treatment of acromegaly.

Literature Cited

1. Ontjes, D. A., Ney, R. L. 1976. Pituitary tumors. *Ca* 26:330–50
2. Glick, S. M. 1975. In *Endocrinology and Diabetes,* ed. L. J. Kryston, R. A. Shaw, pp. 59–67. New York: Grune & Stratton
3. Cryer, P. E., Daughaday, W. H., Coxe, W. S., Avioli, L. V. 1975. Diagnosis and therapy of acromegaly. *Arch. Intern. Med.* 135:338–43
4. Linfoot, J. A., Chong, C. Y., Lawrence, J. H., Born, J. L., Tobias, C. A., Lyman, J. 1975. In *Hormonal Proteins and Peptides,* ed. C. H. Li, pp. 191–246. New York: Academic
5. Editorial. 1974. Assessment and management of acromegaly. *Br. Med. J.* 4:549–50
6. Wright, A. D., Hill, D. M., Lowy, C., Fraser, T. R. 1970. Mortality in acromegaly. *Q. J. Med.* 39:1–16
7. Goldfine, I. D., Lawrence, A. M. 1972. Hypopituitarism in acromegaly. *Arch. Intern. Med.* 130: 720–23
8. Roth, J., Gorden, P., Brace, K. 1970. Efficacy of conventional pituitary irradiation in acromegaly. *N. Engl. J. Med.* 282:1385–91
9. Lawrence, A. M., Pinsky, S. M., Goldfine, I. D. 1971. Conventional radiation therapy in acromegaly. *Arch. Intern. Med.* 128:369–77
10. Aloia, J. F., Roginsky, M. S., Archambeau, J. O. 1974. Pituitary irradiation in acromegaly. *Am. J. Med. Sci.* 267:81–87
11. Atkinson, R. L., Becker, D. P., Martins, A. N., Schaaf, M., Dimond, R. C., Wartofsky, L., Earll, J. M. 1975. Acromegaly treatment by transsphenoidal microsurgery. *J. Am. Med. Assoc.* 233:1279–83
12. Williams, R. A., Jacobs, H. S., Kurtz, A. B., Millar, J. G. B., Oakley, N. W., Spathis, G. S., Sulway, M. J., Nabarro, J. D. N. 1975. The treatment of acromegaly with special reference to trans-sphenoidal hypophysectomy. *Q. J. Med.* 44:79–98
13. Kinnman, J. 1976. The prognosis is acromegaly treated by transanthrosphenoidal operation. *Acta Oto-laryngol.* 82:420–30
14. Merimee, T. J., Rabinowitz, D. 1974. *Isolated Human Growth Hormone Deficiency and Related Disorders.* New York: Intercont. Med. Books. 63 pp.
15. Martin, J. B. 1973. Neural reg-

ulation of growth hormone secretion. *N. Engl. J. Med.* 285:1384–93

16. Müller, E. E. 1973. Nervous control of growth hormone secretion. *Neuroendocrinology* 11:338–69
17. Taylor, A. L. 1976. In *Hormones and Cancer,* ed. K. K. N. Charyulu, A. Sudarsanam, pp. 1–8. New York: Symposia Specialists
18. Brazeau, P., Vale, W., Burgus, R., Ling, N., Butcher, M., Rivier, J., Guillemin, R. 1973. Hypothalamic polypeptide that inhibits the secretion of immunoreactive pituitary growth hormone. *Science* 179:77–79
19. Roth, J., Glick, S. M., Cuatrecasas, P., Hollander, C. S. 1967. Acromegaly and other disorders of growth hormone secretion. *Ann. Intern. Med.* 66:760–88
20. Beck, P., Schalch, D. S., Parker, M. L., Kipnis, D. M., Daughaday, W. H. 1965. Correlative studies of growth hormone and insulin plasma concentrations with metabolic abnormalities in acromegaly. *J. Lab. Clin. Med.* 66:366–79
21. Cryer, P. E., Daughaday, W. H. 1969. Regulation of growth hormone secretion in acromegaly. *J. Clin. Endocrinol.* 29:386–93
22. Lawrence, A. M., Goldfine, I. D., Kirsteins, L. 1970. Growth hormone dynamics in acromegaly. *J. Clin. Endocrinol.* 31:239–47
23. Liuzzi, A., Chiodini, P. G., Botalla, L., Cremascoli, G., Silverstrini, F. 1972. Inhibitory effect of L-dopa on GH release in acromegalic patients. *J. Clin. Endocrinol. Metab.* 35:941–43
24. Irie, M., Tsushima, T. 1972. Increase of serum growth hormone concentration following thyrotropin-releasing hormone injection in patients with acromegaly or gigantism. *J. Clin. Endocrinol. Metab.* 35:97–100
25. Faglia, G., Beck-Peccoz, P., Travaglini, P., Paracchi, A., Spada, A., Lewin, A. 1973. Elevations in plasma growth hormone concentration after luteinizing hormone–releasing hormone (LRH) in patients with active acromegaly. *J. Clin. Endocrinol. Metab.* 37:338–40
26. Hagen, T. C., Lawrence, A. M., Kirsteins, L. 1972. Autoregulation of growth hormone secretion in normal subjects. *Metabolism* 21:603–10
27. Hagen, T. C., Lawrence, A. M., Kirsteins, L. 1971. *In vitro* release of monkey pituitary growth hormone by acromegalic plasma. *J. Clin. Endocrinol.* 33:448–51
28. Schwartz, E., Echemendia, E., Schiffer, M., Panariello, V. A. 1969. Mechanism of estrogenic action in acromegaly. *J. Clin. Invest.* 48:260–70
29. Simon, S., Schiffer, M., Glick, S. M., Schwartz, E. 1967. Effect of medroxyprogesterone acetate upon stimulated release of growth hormone in men. *J. Clin. Endocrinol.* 27:1633–36
30. Lawrence, A. M., Kirsteins, L. 1970. Progestins in the medical management of active acromegaly. *J. Clin. Endocrinol.* 30:646–52
31. Malarkey, W. B., Daughaday, W. H. 1971. Variable response of plasma GH in acromegalic patients treated with medroxyprogesterone acetate. *J. Clin. Endocrinol.* 33:424–31
32. Camanni, F., Massara, F., Fassio, V., Molinatti, G. M., Müller, E. E. 1975. Effect of five dopaminergic drugs on plasma growth hormone levels in acromegalic subjects. *Neuroendocrinology* 19:227–40
33. Sachdev, Y., Tunbridge, W. M. G., Weightman, D. R., Gomez-Pan, A., Duns, A., Hall, R., Goolamali, S. K. 1975. Bromocriptine therapy in acromegaly. *Lancet* 2:1164–68
34. Camanni, F., Massara, F., Belforte, L., Molinatti, G. M. 1975. Changes in plasma growth hormone levels in normal and acromegalic subjects following administration of 2-bromo-α-ergocryptine. *J. Clin. Endocrinol. Metab.* 40:363–66
35. Benker, G., Zäh, W., Hackenberg, K., Hamburger, B., Günnewig, H., Reinwein, D. 1976. Long-term treatment of acromegaly with bromocryptine: Postprandial HGH levels and response to TRH and glucose administration. *Horm. Metab. Res.* 8:291–95
36. Summers, V. K., Hipkin, L. J., Diver, M. J., Davis, J. C. 1975. Treatment of acromegaly with bromocryptine. *J. Clin. Endocrinol. Metab.* 40:904–6
37. Cassar, J., Mashiter, K., Joplin, G. F. 1977. Bromocriptine treatment of acromegaly. *Metabolism* 26:539–46
38. Wass, J. A. H., Thorner, M. O.,

Morris, D. V., Rees, L. H., Mason, A. S., Jones, A. E., Besser, G. M. 1977. Long-term treatment of acromegaly with bromocriptine. *Br. Med. J.* 1:875–78

39. Boyd, A. E. III, Reichlin, S. 1975. Galactorrhea-amenorrhea, bromoergocryptine, and the dopamine receptor. *N. Engl. J. Med.* 293:451–52
40. Kolodny, H. D., Sherman, L., Singh, A., Kim, S., Benjamin, F. 1971. Acromegaly treated with chlorpromazine. *N. Engl. J. Med.* 284:819–22
41. Dimond, R. C., Brammer, S. R., Atkinson, R. L. Jr., Howard, W. J., Earll, J. M. 1973. Chlorpromazine treatment and growth hormone secretory responses in acromegaly. *J. Clin. Endocrinol. Metab.* 36:1189–95
42. Nakagawa, K., Mashimo, K. 1973. Suppressibility of plasma growth hormone levels in acromegaly with dexamethasone and phentolamine. *J. Clin. Endocrinol. Metab.* 37:238–46
43. Sulaiman, W. R., Johnson, R. H., Rennie, M. J. 1975. Hormonal and metabolic studies of the action of fenfluramine in normal man, obesity, acromegaly and diabetes melitus. *Postgrad. Med. J.* 51:125–31
44. Feldman, J. M., Plonk, J. W., Bivens, C. H. 1976. Inhibitory effect of serotonin antagonists on growth hormone release in acromegalic patients. *Clin. Endocrinol.* 5:71–78
45. Delitala, G., Masala, A., Alagna, S., Devilla, L., Lotti, G. 1976. Growth hormone and prolactin release in acromegalic patients following metergoline administration. *J. Clin. Endocrinol. Metab.* 43:1382–86
46. Besser, G. M., Mortimer, C. H., Carr, D., Schally, A. V., Coy, D. H., Evered, D., Kastin, A. J., Tunbridge, W. M. G., Thorner, M. O., Hall, R. 1974. Growth hormone release inhibiting hormone in acromegaly. *Br. Med. J.* 1:352–55
47. Brown, M., Rivier, J., Vale, W. 1977. Somatostatin: Analogs with selected biological activities. *Science* 196:1467–69
48. Takahara, J., Hosogi, H., Yamauchi, J., Ogawa, N., Ofuji, T. 1975. Suppression of serum growth hormone levels by glucagon in patients with active acromegaly. *Endocrinol. Jpn.* 22:581–84
49. Tamburrano, G., Tamburrano, S., Gambardella, S., Andreani, D. 1976. Effects of alcohol on growth hormone secretion in acromegaly. *J. Clin. Endocrinol. Metab.* 42:193–96

Ann. Rev. Med. 1978. 29:417–26
Copyright © 1978 by Annual Reviews Inc. All rights reserved

SLOW-CHANNEL DEPOLARIZATION: MECHANISM AND CONTROL OF ARRHYTHMIAS[1]

◆7289

John C. Bailey, M.D., Victor Elharrar, Ph.D., and Douglas P. Zipes, M.D.
Krannert Institute of Cardiology, Department of Medicine, Indiana University School of Medicine, and Veterans Administration Hospital, Indianapolis, Indiana 46202

INTRODUCTION

When cardiac cells are sufficiently stimulated, they undergo a sequence of transmembrane voltage changes known as a transmembrane action potential. The configuration of the action potential is due, in part, to the flow of depolarizing and repolarizing ionic currents through various channels in the cell membrane. One such current is a secondary inward depolarizing current that flows through a slow channel. This current is probably carried primarily by calcium ions, and voltage changes due solely to this current are known as slow-channel depolarizations or slow responses. This paper describes the electrophysiology of slow-channel depolarizations and reviews some of the possible physiological, pathophysiological, and pharmacological implications of the responses.

ELECTROPHYSIOLOGY

The electrical diastole of typical cardiac atrial, His-Purkinje, or ventricular myocardial cells is characterized by a steady transmembrane voltage gradient of approximately 90 mV; the interior of the cell is negative compared to the extracellular compartment (1). Suprathreshold stimulation of these cells produces a rapid change of the transmembrane voltage gradient, so that the interior of the cell becomes positively charged compared to the exterior. This rapid (500–1000 V/sec) depolarization is due to a sudden increase in the permeability

---

[1] Supported in part by the Herman C. Krannert Fund, and by Grants HL-06308, HL-07182, HL-18795, and HL-18538 from the National Heart, Lung, and Blood Institute of the National Institutes of Health, US Public Health Service, and the American Heart Association, Indiana Affiliate, Inc.

0066-4219/78/0401-0417$01.00

of the cell membrane to the sodium ion (2). The sodium concentration in the extracellular space is approximately 145 mEq/l, whereas the intracellular concentration of the ion is approximately 5 mEq/l. This steep concentration gradient therefore facilitates the rapid inward movement of sodium ions. The magnitude of the transmembrane potential gradient (−90 mV), the amplitude of depolarization (~120 mV), and the rate of depolarization (500–1000 V/sec) account for the rapid conduction velocity (1–4 m/sec) recorded normally in these cardiac tissues (3). This rapid inward sodium current through a fast channel accounts for the bulk of the voltage change during phase 0 of the transmembrane action potential. Typical fast-channel–dependent action potentials are illustrated in Figures 1A and 2A. The terminal portion of phase 0 as well as the plateau (phase 2) of the action potential is due, at least in part, to a much slower and smaller secondary inward current through a separate slow channel (4). The slow inward current is activated at a low transmembrane potential (−40 mV) (5–9), is relatively rapidly activated but slowly inactivated (10–12), is not very sensitive to external sodium ion concentration [In fact it operates when extracellular sodium concentration is zero (13–15) or when the fast sodium channel is inactivated.] (5–11, 16), but is exquisitely sensitive to external calcium ion concen-

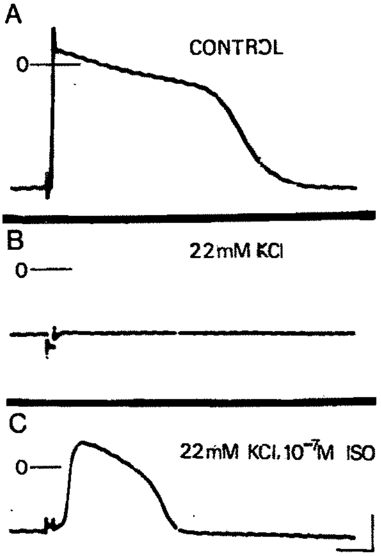


Figure 1 Generation of slow-current depolarization. Panel A: Control transmembrane action potential recorded from canine cardiac Purkinje fiber. Note high resting potential (−90mV), action-potential amplitude (115 mV), and rapid depolarization velocity. Panel B: Addition of 22 mM KCl depolarized the membrane to about −40 mV and produced inexcitability, even with increased stimulus intensity. Panel C: Addition of 10^{-7} M isoproterenol restored excitability, which gave rise to slow-response electrical activity (see text). Line marked 0 indicates zero transmembrane potential gradient. Calibrations: horizontal bar = 50 msec, vertical bar = 25 mV.

tration (11, 12, 16, 17). This slow inward current is unaffected by specific fast-channel blockers (6, 8, 11) but is inhibited by ions and compounds that do not affect the fast inward current (11, 12, 16, 17).

The rapid sodium-dependent and the slow calcium-dependent depolarizing currents differ considerably in their time and voltage dependence. Therefore cardiac cells can be manipulated in various ways so that transmembrane voltage changes consequent only to the slow inward current may be recorded. These potential changes i.e. slow-channel depolarizations or slow responses are illustrated in Figures 1C, 2B, and 2C.

The electrophysiological characteristics of slow responses include low transmembrane resting potential (−40 mV), low action-potential amplitude (35–50 mV), very slow rate of depolarization (~10 V/sec), very slow propagation velocity (<< 0.1 m/sec), graded response to applied stimuli, refractoriness exceeding action-potential duration, and very low safety factor for propagation (18). Moreover, this slow inward calcium current is augmented by catecholamines (7, 19). Also, although the normal diastolic pacemaker current is inactivated in the voltage range where slow-current–dependent potentials are generated, these low-amplitude slow responses are also capable of spontaneous, sustained rhythmic activity by another unknown mechanism (20).

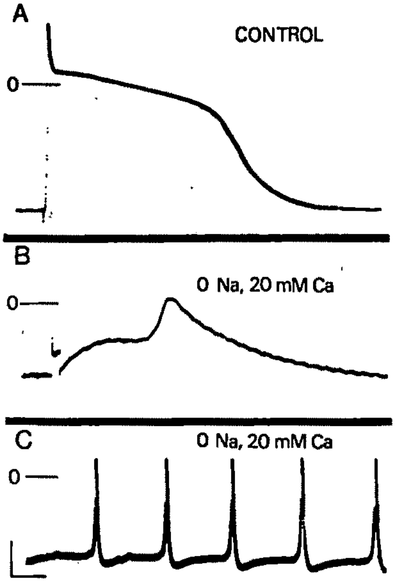


Figure 2 Generation of another type of slow-current depolarization. Panel A: Control Purkinje transmembrane action potential. Panel B: Slow-current depolarization produced in a solution containing no Na$^+$ and 20 mM Ca^{++}. Panel C: Automaticity in a fiber demonstrating only slow-current depolarizations. Note the slow decline of transmembrane potential between action potentials. Line marked 0 and calibrations are as in Figure 1, except horizontal line for Panel C is 1 sec.

EXPERIMENTAL GENERATION OF THE SLOW RESPONSE

A frequently used method of generating slow responses in vitro includes elevation of extracellular potassium ion concentration, which causes a decline of transmembrane resting voltage from −90 mV to approximately −40 mV (Figure 1*b*). At this lower transmembrane potential the fast inward sodium current is inactivated. Exposure of these depolarized fibers to appropriate concentrations of catecholamines will enhance the slow inward calcium current and generate a slowly rising, low-amplitude, easily extinguished potential that is independent of sodium influx through the fast channel (Figure 1C) (21). Exposure of cardiac fibers to a sodium-free, calcium-rich medium in vitro permits the generation of slow responses (Figure 2B) that often demonstrate the property of automaticity (Figure 2C) (22), i.e. the ability to generate electrical potentials without external stimulation. This form of slow-response automaticity may be due to an inward calcium current during electrical diastole (23).

SIGNIFICANCE OF THE SLOW INWARD CALCIUM CURRENT

Excitation-Contraction Coupling

It has been known for many years that cardiac contraction does not occur in the absence of calcium ions in the extracellular space. Recent studies (24) demonstrate that significant cardiac muscle contractions develop at a transmembrane potential (−30 mV) similar to the voltage threshold for the slow inward current. Peak tension is developed only after a series of action potentials (staircase phenomenon), although the magnitude of the inward calcium current does not increase with each successive action potential. It appears, therefore, that the transmembrane influx of calcium with each action potential releases intracellular calcium stores that interact with the contractile proteins to cause the contractile event. In any case it seems clear that excitation-contraction coupling is dependent on the slow inward calcium current.

Autonomic Interactions

It has been demonstrated (4, 25) that catecholamines augment the inward calcium current and, *pari passu*, exert a positive inotropic response in cardiac tissue. Watanabe & Besch (26) showed that the positive inotropic effects of isoproterenol were significantly blunted when the catecholamine was administered simultaneously with acetylcholine. Other studies have shown that cholinergic activity increases electrical stability of the ventricles only when adrenergic tone is increased (27). Recent experiments from our laboratories (28) demonstrate that the slow response generated by catecholamine stimulation is extinguished by the muscarinic effects of acetylcholine. The slow response elicited in a sodium-free solution (i.e. independent of beta-adrenergic stimulation) is unaffected by acetylcholine. These data may provide a cellular electrophysiological basis for the antiarrhythmic effect of cholinergic stimulation in the setting of myocardial ischemia when cardiac sympathetic tone is high (27).

Cardiac Arrhythmias in Myocardial Infarction

Cardiac arrhythimias are most likely caused by disorders of impulse formation, impulse propagation, or a combination of these abnormalities. Soon after acute myocardial infarction, extracellular potassium ion concentration in the ischemic zone probably increases (29) and may be responsible for the release of local endogenous catecholamines. The similarities between the experimental conditions required to generate slow responses in vitro noted above and the pathophysiological milieu that may accompany acute myocardial infarction have led to the supposition that slow-channel–dependent electrical responses may be responsible for the arrhythmias accompanying myocardial infarction (18, 30, 31). For example, specialized intraventricular conducting fibers might, under the conditions just described, depolarize abnormally and undergo spontaneous diastolic depolarization in a voltage range where the normal ionic conductance changes responsible for the automatic process in such fibers cannot operate. This circumstance would give rise to an ectopic ventricular rhythm due to automaticity in fibers manifesting electrical potentials characteristic only of slow responses.

Reentrant arrhythmias due to unidirectional conduction block have been postulated for many years (32). According to this hypothesis, illustrated in Figure 3, an impulse that ultimately produces normal activation of, for example, the ventricles fails to penetrate a small portion of the intraventricular conducting system. Activation of the ventricle distal to the site of orthograde block occurs via another pathway. That portion of conducting tissue blocked to orthograde conduction, however, permits retrograde penetration, which allows eventual reexcitation of the entire ventricle. Certain conditions must exist before such reentrant reexcitation can occur. Specifically, the length of the reentrant pathway must be equal to or exceed the product of the refractory period duration and

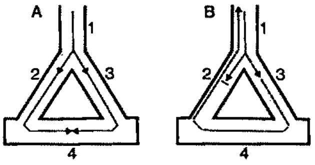


Figure 3 A simplified diagram of circuitous reentry. Panel A: Normal activation sequence. The cardiac impulse activates a penultimate pathway 1, then bifurcates to engage pathways 2 and 3. The two wavefronts activate pathway 4, collide, and are extinguished. Panel B: Activation sequence giving rise to reentry. The impulse engages pathway 1, bifurcates, propagates with delay along pathway 3, but is blocked in pathway 2. Pathway 4 is eventually activated, and, since orthograde conduction in pathway 2 had failed due to unidirectional block, pathways 2 then 1 may be engaged retrogradely, causing eventual reexcitation of the heart.

the conduction velocity. Electrical activity due to a fast inward sodium current propagates at rates of 1–3 m/sec with refractory period durations of 200–500 msec. A pathway of at least 20 cm would be required to sustain a reentrant rhythm, given a conduction velocity of 1 m/sec and a refractory period of 200 msec. Clearly no such pathway exists in the ventricles. On the other hand, cardiac tissue manifesting only slow-response electrical activity might easily permit reentrant rhythms, since the propagation speed of slow response is reduced to as low as 0.05 m/sec. Discovery of slow cardiac responses thus has added credence to the notion that ectopic ventricular rhythms, especially those that are coupled to previous, ostensibly normal, impulses, may be due to a reentry mechanism (33, 34).

Experimental demonstration that slow-response electrical activity accounts for ventricular arrhythmias that accompany myocardial infarction is lacking, although conclusive evidence for the operation of any particular mechanism in the genesis or perpetuation of a given arrhythmia may be impossible to obtain. Several recent reports have described the cellular electrophysiological alterations of infarcted myocardium in vitro using microelectrode techniques. Friedman, et al (35, 36) studied such preparations 16–24 hr after coronary artery ligation in the dog. They determined that no electrical activity could be recorded from ventricular muscle, but that subendocardial Purkinje fibers within the infarcted zone frequently survived. Transmembrane potentials from these cells were characterized by reduced transmembrane resting potential, lowered action-potential amplitude, reduced maximum rate of depolarization, spontaneous diastolic depolarization, and remarkably prolonged action-potential durations. Lazzara et al (37) noted similar abnormalities in ventricular muscle and subendocardinal Purkinje fibers of dogs 24 hr after experimental myocardial infarction. They reported, however, that after the preparations had been superfused in vitro with oxygenated Tyrode's solution for two hours, the electrophysiological parameters of these Purkinje cells reverted toward normal, except that prolongation of action-potential duration persisted. More recently, Friedman and co-workers (38) reported that Purkinje fibers surviving three days to seven weeks after experimental myocardial infarction persistently demonstrated reduced transmembrane resting potential and action-potential amplitude, decreased rate of depolarization, and reduced conduction velocity. In these microelectrode studies of myocardial infarction, although fibers in the infarction zone demonstrated "depressed" electrical activity, typical slow responses were not routinely recorded, and a preponderance of the reported cells did not satisfy the criteria for responses due to an inward current through the slow channel. In fact, recent preliminary reports (39, 40) implicate fast-sodium-channel dependence of depressed action potentials in a zone of myocardial infarction. The evidence, therefore, favoring slow-response electrical activity as a cause of clinical arrhythmias during myocardial infarction is indirect and, at best, circumstantial. Further stringent quantitation of cellular electrophysiological alterations accompanying myocardial infarction, documentation of the effects of specific slow-channel blockers and enhancers on electrical activity during infarction, and correlation of these studies with the presence or absence of arrhythmias are needed before definitive statements regarding the mechanism for arrhythmias in myocardial infarction can be made.

Sinoatrial and Atrioventricular Nodal Electrical Activity

Transmembrane action potentials recorded from single cells of the sinoatrial (SA) node closely resemble morphologically the slow response generated in a sodium-free, calcium-rich medium (Figure 2C). Action potentials from the nodal region of the atrioventricular (AV) junction also resemble slow responses (e.g. Figures 1C and 2B), although automaticity is not a property of the true nodal region of the AV junction. Moreover the SA and AV nodes behave electrophysiologically like cardiac Purkinje fibers demonstrating slow responses. Thus conduction speed in these nodal regions is quite slow, conduction block is commonplace, and graded electrical responses are frequent (41). Furthermore refractoriness in these tissues outlasts action-potential duration (41). These similarities between nodal and slow-response electrical activity suggest that depolarization in SA and AV nodal cells may depend on an inward current through the slow channel. This hypothesis is strengthened considerably by the observations that nodal potentials are depressed by agents that block calcium current (17, 42, 43) but are unaffected by drugs that selectively block the fast inward sodium current (43, 44). Of particular interest has been the demonstration that verapamil, an antiarrhythmic agent in clinical use in Europe, blocks the slow channel in pharmacologically relevant concentrations. In doses that do not significantly affect action potentials of the His-Purkinje system, this drug produces a decrease in the action-potential amplitude and upstroke velocity of SA and AV nodal cells, and reduces the rate of spontaneous SA node discharge (42). In isolated AV node preparations, verapamil slows AV transmission and prolongs AV node refractoriness. AV node reentry, one likely mechanism for reciprocating tachycardia (45), is prevented by verapamil, because this drug, by prolonging AV node refractoriness, blocks rather than delays transmission of premature impulses (42). Clinical reports indicate that verapamil is effective in the treatment of supraventricular tachycardias (46). Thus it is tempting to suggest, in view of this drug's slow-channel blocking property and its actions on SA and AV nodal electrical activity (47), that depolarization in these tissues is dependent on the slow inward current. At the present time this suggestion remains an attractive hypothesis.

SUMMARY

The secondary inward current that flows through the slow channel is probably carried primarily by calcium ions. This current is responsible, in part, for the plateau phase of the cardiac action potential. Inward calcium current through the slow channel is essential to excitation-contraction coupling, and enhancement of this current exerts a positive inotropic effect. Transmembrane electrical potentials generated as a consequence of this slow inward current, so-called

slow-channel depolarizations or slow responses, have been useful models in recent studies of cardiac autonomic interactions. The significance of the slow response in ventricular arrhythmias remains controversial and awaits more definitive experiments. The role of slow-channel depolarization as a basis for SA and AV nodal electrical activity is suggested by the electrophysiological similarities between these tissues and slow responses produced in vitro. This hypothesis is supported by the responses of these nodal tissues to interventions that augment or impede the slow inward current. More direct evidence in support of this notion may not be obtainable in the near future, since the critical voltage-clamp studies of ionic currents in SA and AV nodes are not technically feasible at this time.

Literature Cited

1. Hoffman, B. F., Cranefield, P. F. 1960. *Electrophysiology of the Heart*, pp. 6–10. New York: McGraw-Hill. 323 pp.
2. Draper, M. H., Weidman, S. 1951. Cardiac resting and action potentials recorded with an intracellular electrode. *J. Physiol. (London)* 115:74–94
3. Hoffman, B. F., Cranefield, P. F. 1960. See Ref. 1, pp. 259–60
4. Reuter, H. 1967. The dependence of slow inward current in Purkinje fibers on the extracellular calcium concentration. *J. Physiol. (London)* 192:479–92
5. Engstfled, G., Antoni, H., Fleckenstein, A. 1961. Die restitution der errengungsfortleitung and kontraktionskraft des K^+-gelahmten frosch und saugertiermyokards durch adrenalin. *Pfluegers Arch.* 273:145–61
6. Carmeliet, E., Vereecke, J. 1969. Adrenalin and the plateau phase of the cardiac action potential. *Pfluegers Arch.* 313:300–15
7. Pappano, A. J. 1970. Calcium-dependent action potentials produced by catecholamines in guinea pig atrial muscle fibers depolarized by potassium. *Circ. Res.* 27:379–90
8. Shigenobu, K., Sperelakis, N. 1972. Calcium current channels induced by catecholamines in chick embryonic hearts whose fast sodium channels are blocked by tetrodotoxin or elevated potassium. *Circ. Res.* 31:932–52
9. Thyrum, P. 1974. Inotropic stimuli and systolic transmembrane calcium flow in depolarized guinea pig atria. *J. Pharmacol. Exp. Ther.* 188:166–79
10. Fozzard, H. A., Gibbons, W. R. 1973. Action potential and contraction of heart muscle. *Am. J. Cardiol.* 31:182–92
11. Rougier, O., Vassort, G., Garnier, D., Gargouil, Y. M., Coraboeuf, E. 1969. Existence and role of slow inward current during the frog atrial action potential. *Pfluegers Arch.* 308:91–110
12. Reuter, H. 1968. Slow inactivation of currents in cardiac Purkinje fibers. *J. Physiol. (London)* 197:233–53
13. Aronson, R. S., Cranefield, P. F. 1973. Electrical activity of canine cardiac Purkinje fibers in sodium-free, calcium-rich solutions. *J. Gen. Physiol.* 61:786–808
14. Delahayes, J. F. 1972. Electrical responses of cardiac muscle in Na-free high-Ca solution. *Experientia* 28:1054–55
15. Verdonick, F., Busselen, P., Carmeliet, E. 1972. Ca-action potentials and contractions of heart muscle in Na-free solutions: Influence of caffeine. *Arch. Int. Physiol. Biochim.* 80:167–69
16. Reuter, H. 1967. Dependence of slow inward current in Purkinje fibers on the extracellular calcium-concentration. *J. Physiol. (London)* 192:479–92
17. Kolhardt, M., Bauer, B., Krause, H., Fleckenstein, A. 1972. Differentiation of the transmembrane Na and Ca channels in mammlian cardiac fibers by the use of specific inhibitors. *Pfluegers Arch.* 335:309–22
18. Zipes, D. P., Besch, H. R. Jr., Watanabe, A. M. 1975. Role of the slow current in cardiac electrophysiology. *Circulation* 51:761–66

19. Vassort, G., Rougier, O., Garnier, D., Sauviat, M. P., Coraboeuf, E., Gargouil, Y. M. 1969. Effects of adrenaline on membrane inward currents during the cardiac action potential. *Pfluegers Arch.* 309:70–81
20. Aronson, R. S., Cranefield, P. F. 1973. The effect of resting potential on the electrical activity of canine cardiac Purkinje fibers exposed to Na-free solution or to ouabain. *Pfluegers Arch.* 347:101–16
21. Engstfeld, G., Antoni, H., Fleckenstein, A., 1961. Die restitution der errengungsfortleitung und kontraktionskraft des K$^+$-gelähmten frosch- und saugetiermyokards durch adrenalin. *Pfluegers Arch.* 273:145–63
22. Aronson, R. S., Cranefield, P. F. 1974. The effect of resting potential on the electrical activity of canine cardiac Purkinje fibers exposed to a Na-free solution or to ouabain. *Pfluegers Arch.* 273:101–16
23. Cranefield, P. F., Aronson, R. S., Wit, A. L. 1974. Effect of verapamil on the normal action potential and on a calcium-dependent slow response of canine cardiac Purkinje fibers. *Circ. Res.* 34:204–13
24. Beeler, G. W., Reuter, H. 1970. The relation between membrane potential, membrane currents and activation of contraction in ventricular myocardial fibres. *J. Physiol.* 207:211–29
25. Rougier, O., Vassort, G., Garnier, D., Gargouil, Y, M., Coraboeuf, E. 1969. Existence and role of a slow inward current during the frog atrial action potential. *Pfluegers Arch.* 308:91–110
26. Watanabe, A. M., Besch, H. R. Jr. 1975. Interaction between cyclic adenosine monophosphate and cyclic guanosine monophosphate in guinea pig ventricular myocardium. *Circ. Res.* 37:309–17
27. Kent, K. M., Epstein, S. E. 1976. Neural basis for the genesis and control of arrhythmias associated with myocardial infarction. *Cardiology* 61:61–74
28. Bailey, J. C., Lathrop, D. A., Besch, H. R. Jr., Watanabe, A. M. 1977. *Clin. Res.* 25(3):206 (Abstr.)
29. Harris, A. S., Bisteni, A., Russell, R. A., Brigham, J. C., Firestone, J. E. 1954. Excitatory factors in ventricular tachycardia resulting from myocardial ischemia: Potassium: a major excitant. *Science* 119:200–3
30. Wit, A. L., Friedman, P. L. 1975. Basis for ventricular arrhythmias accompanying myocardial infarction. *Arch. Intern. Med.* 135:459–72
31. Cranefield, P. F. 1975. *The Conduction of the Cardiac Impulse,* p. 308. Mount Kisco, New York: Futura
32. Mines, G. R. 1913. On dynamic equilibrium in the heart. *J. Physiol.* 46:349–383
33. Wit, A. L., Hoffman, B. F., Cranefield, P. F. 1972. Slow conduction and reentry in the ventricular conducting system: I. Return extrasystole in canine Purkinje fibers. *Circ. Res.* 30:1–10
34. Wit, A. L., Cranefield, P. F., Hoffman, B. F. 1972. Slow conduction and reentry in the ventricular conducting system: II. Single and sustained circus movement in networks of canine and bovine Purkinje fibers. *Circ. Res.* 30:11–22
35. Friedman, P. L., Stewart, J. R., Fenoglio, J. J. Jr., Wit, A. L. 1973. Survival of subendocardial Purkinje fibers after extensive myocardial infarction in dogs: In vitro and in vivo correlations. *Circ. Res.* 33:597–611
36. Friedman, P. L., Stewart, J. R., Wit, A. L. 1973. Spontaneous and induced cardiac arrhythmias in subendocardial Purkinje fibers surviving extensive myocardial infarction in dogs. *Circ. Res.* 33:612–26
37. Lazzara, R., El-Sherif, N., Scherlag, B. J. 1973. Electrophysiological properties of canine Purkinje cells in one-day-old myocardial infarction. Circ. Res. 33:722–34
38. Friedman, P. L., Fenoglio, J. J. Jr., Wit, A. L. 1975. The time course for reversal of electrophysiological and ultrastructure abnormalities in subendocardial Purkinje fibers surviving in regions of extensive myocardial infarction. *Circ. Res.* 36:127–44
39. Brennan, F. J., Wit, A. L., Cranefield, P. F. 1975. Circulation 51 & 52:Suppl. II, p. 130
40. Lazzara, R., El-Sherif, N., Befler, B., Scherlag, B. 1975. *Circulation* 51 & 52:Suppl. II, p. 330
41. Hoffman, B. F., Cranefield, P. F. 1960. See Ref. 1, pp. 118–20, 158
42. Wit, A. L., Cranefield, P. F. 1974.

Effect of verapamil on the sinoatrial and atrioventricular nodes of the rabbit and the mechanism by which it arrests reentrant atrioventricular nodal tachycardia. *Circ. Res.* 35: 413–25

43. Zipes, D. P., Mendez, C. 1973. Action of manganese ion and tetrodotoxin on atrioventricular nodal transmembrane potentials in isolated rabbit hearts. *Circ. Res.* 32:447–54
44. Yamagishi, S., Sano, T. 1966. Effect of tetrodotoxin on the pacemaker action potential of the sinus node. *Proc. Jpn. Acad.* 42:1194–96
45. Bigger, J. T., Goldreyer, B. N. 1970. Mechanism of supraventricular tachycardia. *Circulation* 42: 673–88
46. Schamroth, L., Krikler, D. M., Garrett, C. 1972. Immediate effects of intravenous verapamil in cardiac arrhythmias. *Br. Med. J.* 1:660–62

REFERENCE ADDED IN PROOF

47. Zipes, D. P., Fischer, J. C. 1974. Effects of agents which inhibit the slow channel on sinus node automaticity and atrioventricular conduction in the dog. *Circ. Res.* 34:124–92

Ann. Rev. Med. 1978. 29:427–57
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CONTRACTILE PROTEINS IN CELL STRUCTURE AND FUNCTION

◆7290

Thomas P. Stossel, M.D.[1]
Medical Oncology Unit, Massachusetts General Hospital, and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114

INTRODUCTION

Within the past decade, the attention of biologists has turned sharply to the "cell sap," the bland matrix between more prepossessing cellular organelles. In light of recent progress in the understanding of cytoplasm, W.H. Lewis's assessment in 1939 was amazingly accurate:

> The cytoplasm is a colloid with extremely large and asymmetrical molecules which presumably form long fibers on gelation. Since most colloidal gels automatically exert contractile tension, it is probable that gelled cytoplasm acts likewise. The assumption that active cells possess a superficial plasmagel layer which exerts contractile tension offers a key for an explanation of changes of form of locomotion and the division of cells when coupled with local and general alterations of the viscosity of the cytoplasm and plasmagel layer. (1)

Aeriel Loewy made the important connection in 1952 that extracts prepared from a slime mold, *Physarum polycephalum*, behaved like mixtures of the then recently discovered major proteins of skeletal muscle, actin and myosin (2). However, the now widespread appreciation of "contractile" proteins awaited a number of events:

1. The isolation of proteins from "nonmuscle" cells and tissues in sufficient yield and purity to be characterized as actin, myosin, and other muscle proteins was necessary. This accomplishment began in the late 1960s and remains the most important foundation of this new understanding of cell structure and function. This endeavor is somewhat difficult to undertake and is slow to reward. Therefore, other developments played a greater role in the current popularity of this field.

2. The development of techniques for identifying contractile proteins in nonmuscle cells by morphologic techniques enabled the microscopist to recognize

---

[1] Supported by USPHS Grant HL 19429 and a grant from the Edwin S. Webster Foundation.

0066-4219/78/0401-0427$01.00

contractile proteins in many cells. One method, introduced in 1969 (3), took advantage of the specific appearance, as viewed in the electron microscope, that heavy meromyosin, a tryptic fragment of skeletal muscle myosin, gave to actin filaments after binding to them. The thin (4–6 nm) actin filaments acquired barbed "arrowhead" projections (4) (Figure 1B). To render the intracellular actin accessible to heavy meromyosin, glycerination of cells was necessary. This procedure sadly distorted normal cell architecture. Despite this limitation, an outpouring of reports employing this technique hinted at the ubiquity of actin in cells, and made fairly certain that one could equate the 4–6-nm cortical "microfilaments" of most cells with actin. The development of a technique that

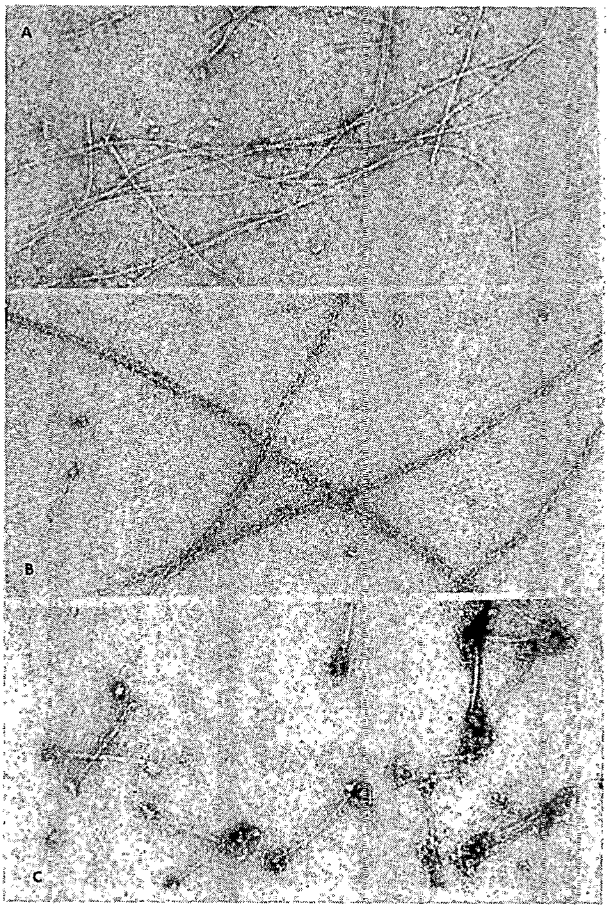


Figure 1 Morphology of contractile proteins visualized in the electron microscope after negative staining with uranyl acetate. A. Actin filaments (F-actin). B. F-actin reacted with skeletal muscle heavy meromyosin and showing barbed unidirectional projections. C. Bipolar myosin filaments.

used fluorescein-conjugated heavy meromyosin to label the actin brought the in situ identification of actin to the level of the light microscope (5). Recently, a related approach, the illumination of cells with fluorochrome-labelled antibodies to various contractile proteins, has become very popular.

3. It was discovered in the late 1960s that a class of fungal products, the cytochalasins, had a variety of spectacular effects on diverse cell movements (6), and in some cases altered the architecture of intracellular filaments (7). One hypothesis for these effects was that the molecular site of action of cytochalasins was on "contractile microfilaments" (7). This idea permitted the cell physiologists easily to relate many cell functions to contractile proteins. Until recently, the evidence for this hypothesis was weak, but experiments with cytochalasins, mainly with a compound called cytochalasin B, brought theories about contractile proteins into a vast spectrum of biological systems.

THE CONTRACTILE PROTEINS

Contractile proteins are proteins that have defined functions in the process whereby striated and smooth muscles generate force. So far, all well-defined nonmuscle [or ‘cytoplasmic‘ (8)] contractile proteins have been found to have counterparts in one or both of these muscle types. In some cases, however, investigators first found contractile proteins in cells other than muscle. A comprehensive description of contractile proteins is beyond the scope of this review, but a brief analysis of the most important structural and functional properties of

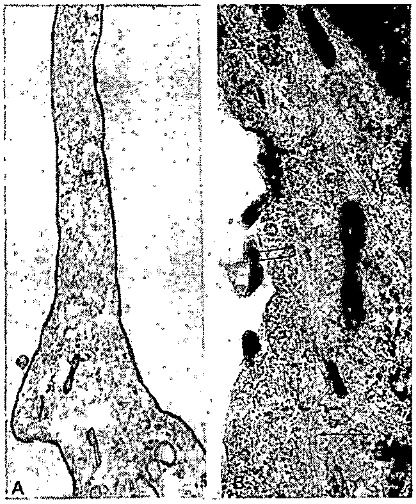


Figure 2 "Cytoskeletal structural elements" as seen with the electron microscope in fixed thin sections of mammalian cells. A. Membrane projection, a filopod or microspike containing cortical microfilaments, which are actin polymers. B. View of cytoplasm showing intermediate 10-nm filaments (*arrow*) and microtubule (*double arrow*).

contractile proteins is necessary. Contractile proteins are important but not the sole elements of cell structure and function. Other "cytoskeletal" elements, primarily the 15-nm hollow microtubules and wavy intermediate 10-nm filaments, are also relevant (3, 9). Contractile proteins are biochemically and morphologically distinct from these other structural proteins and should not be confused with them.[2] (Figure 2).

Actin

Actin is the chief contractile protein. It is, on a molar basis, the most abundant single protein of muscle and of many other cells. The most familiar function of actin is its role as a cable to shorten the striated-muscle contractile unit, or sarcomere (Figure 2), described in more detail below. Equally well-known but much less well-understood is the property of actin to exist in two states: a globular monomer (G-actin) and a double-helical linear polymer (F-actin) (Figure 1A). The polymerized form is the familiar state of actin in skeletal muscle and is now known to be the microfilament of cytoplasm (3). Actin filaments have a definite polarity, shown by the fact that myosin binds to a given helix at an angle determined by the direction of the polarity. The binding of heavy meromyosin to actin displays this polarity, because the arrowhead heavy meromyosin projections from the actin visible in the electron microscope all point in one direction (4). The polymerization of G-actin occurs when the concentration of G-actin monomers in the presence of neutral salts is sufficiently high to create a nucleus for condensation (12). For assembly, the actin monomers must be in a "native" state. ATP and divalent cations bound to actin preserve this native condition, and during polymerization, ATP hydrolysis to ADP usually occurs. Insofar as this hydrolysis is not absolutely required for actin assembly, its role is uncertain (12, 13). Considerable study of the reversible assembly of skeletal-muscle actin has revealed that the kinetics and properties of assembly are extremely complex and depend on many exogenous physical and chemical factors. Until recently, one might have supposed that the reversible assembly of actin was simply a blessing for investigators, since cycles of polymerization provided a simple method for purifying the protein from muscle. Now, as amplified below, it is popular to speculate that polymerization of actin is relevant for all sorts and conditions of cell functions (14). Actin also has other properties

[2] Confusion arises from three sources: (*a*) Tubulin, the subunit protein of microtubules, like actin forms fibroid aggregates involved in cell structure and function. Both actin and tubulin may undergo reversible assembly in vivo and in vitro. (*b*) In some experiments on cell functions, cytochalasins may have effects similar to or synergistic with those of drugs that interfere with the assembly of microtubules, i.e. colchicine and vinca alkaloids. These experimental results are usually difficult to interpret and are presented with vague conclusions concerning "cytoskeletal elements," or "microtubule-microfilament function." (*c*) There have been some claims for structural homologies and biochemical interactions between actin and tubulin (10). The idea that cells' structural systems might interact with one another is appealing, but the specific proposals have either been disputed (11) or not clearly confirmed.

of undetermined relevance. It is an inhibitor of a ubiquitous enzyme, deoxyribonuclease I (15). Actin binds with extraordinary tenacity to this enzyme and, once bound, cannot be dissociated except by strong denaturing solvents (16). Both G- and F-actins bind, but F-actin depolymerizes as it binds the enzyme (17). Actin also binds to enzymes of intermediary metabolism, phosphorylase B, kinase, and aldolase (18, 19).

The actin molecule is remarkably conserved in evolutionary terms. The subunit molecular weight of the actin monomer, about 42,000, is nearly identical for all actins isolated thus far, and limited amino-acid-sequence data indicate very few variations between actins from tissue to tissue or species to species (20). This high degree of molecular monotony may reflect the multiple functions of this protein and its interaction with so many other molecules. The invariance is not absolute, however, inasmuch as subtle differences in net charge of various actins are detectable by isoelectric focussing (21). Differences are also detectable in the molecular weight of actins by electrophoresis on acrylamide slab gels in the presence of urea and sodium dodecylsulfate (22), in peptide mapping (21, 23), and in amino-acid sequence. These techniques found small differences in the structure of skeletal-muscle, brain, and platelet actins and between actins of immature and mature muscle cells (20–23). The presence of isomolecular species of actin in different tissues and during different developmental stages of tissues raises the possibility of using this protein as a genetic or developmental marker in various physiologic or disease states.

Myosin

Myosin is the second most abundant protein of striated muscle but constitutes a much smaller fraction of the protein of nonmuscle cells (Table 1). Myosin in muscle converts chemical energy in ATP into force that shortens the sarcomere, and presumably myosin has a similar role in nonmuscle cells. Almost without exception, myosins are large asymmetrical molecules with complex substructures. Each molecule is composed of a pair of noncovalently linked, large polypeptide, heavy chains of about 220,000 daltons each. These polypeptides have two distinct domains: a globular head region and a helical tail or rod. In addition to the heavy chains, myosins have two or three pairs of low-molecular-weight (15,000–30,000) light chains (24). Under appropriate ionic conditions, including those likely to exist within cells, the tails of the myosin molecules aggregate. The pattern of condensation is usually such that the globular heads project as tufts from both ends of a filamentous aggregate. The resulting "bipolar" myosin filaments, visible in the electron microscope after negative staining, have a central bare shaft with tapered ends from which the heads dangle (Figure 1C). An estimated 28 myosin molecules are in a striated-muscle myosin filament (24a), whereas myosin monomers comprise a filament of platelet myosin (25). The length of myosin filaments depends on the concentration of myosin, the pH, ionic strength, and divalent cation content of the medium, and possibly the origin of the myosin protein.

Table 1 Comparison of actin and myosin contents among striated, smooth, and nonmuscle cells[a]

Cell type	Cell Protein (%)		Actin:Myosin molar ratio
	Actin	Myosin	
Skeletal muscle (rabbit)	23	38	7
Smooth muscle (porcine carotid)	25	8	34
Human platelet	10	1	109
Slime mold (*Dictyostelium Discoideum*)	7	0.5	157

[a] Data from Refs. 100, 160, and 193.

An important feature of myosins is their ATPase activities. The domain of ATP-hydrolyzing activity is the globular head region of the myosin. The conditions under which ATP cleaves to ADP and orthophosphate by myosin are complex and influenced by the origin and state of the myosin, by monovalent and divalent cations, by actin, and by other contractile proteins (26, 27). Native myosins of high purity have active rates of ATP cleavage (over 1 μmole of ATP split per mg of myosin protein per minute at 25°C) when they are in solutions of high ionic strength, and in the presence of calcium but not in the presence of magnesium, which inhibits this ATPase activity. EDTA, which chelates magnesium ions present in ATP, activates the myosin-ATPase activity. The measurement of the EDTA-activated ATPase activity in 0.6 M KCl solution is a useful assay for detecting the presence of myosin and for use during its purification, since most other cellular ATPases are dependent on divalent cations. But note that the ATPase of myosin under these nonphysiologic conditions bears no relevance to the contractile mechanisms to be discussed, and that the calcium-activated ATPase of myosin mentioned above is unrelated to calcium control of contractile activity. The physiologically relevant enzymatic activity is a magnesium ATPase that is activated by actin and controlled by regulatory proteins and, in some systems, calcium.

Despite their considerable similarities in structure and ATPase activity (28), myosins seem much less well-conserved in evolutionary terms than actins. Most antibodies raised against muscle or cytoplasmic myosins show limited cross-reactivity between cell types or between species (29–34). The molecular weight and charge of myosin light chains differ somewhat from tissue to tissue (28), and as discussed below, the mechanism of regulation of the myosin–magnesium ATPase, which is pertinent for movement, varies.

Tropomyosin

Tropomyosin is an asymmetrical molecule of 60,000–70,000 daltons, composed of two noncovalently linked subunits. Tropomyosin is associated with actin of muscle cells (35). Although it exists in nonmuscle cells (36–39), its only known function is that it constitutes part of the protein complex that regulates contractile tension development of striated muscles (40, 41). One tropomyosin molecule binds to seven actin monomers in the groove of the muscle F-actin helix. It forms a ridge that can prevent other proteins, principally myosin heads, from interacting with actin (42, 43) (Figure 3). Actin filaments that have bound tropomyosin have increased rigidity and resistance to chemical depolymerization. Skeletal-muscle tropomyosin differs structurally from its nonmuscle counterparts in several respects; it has slightly larger polypeptide chains and yields different tryptic peptides (38). Antibodies raised against skeletal-muscle tropomyosin react with a variety of nonmuscle cells (44).

Troponins

The troponins are a complex of proteins also associated with actin of certain muscles (42, 43). Three components comprise the troponin complex of muscle: troponin T (38,000 daltons), which binds to actin and to tropomyosin, and thus causes tropomyosin and troponin I (24,000 daltons) to block the interaction between actin and myosin; and troponin C (18,000 daltons), which binds calcium ions. The binding of calcium to troponin C generates by an allosteric mechanism a change in the position of troponin I and tropomyosin. The complex of tropomyosin and the troponins of muscle is termed relaxing or control proteins. Troponin C has structural similarities to various other calcium-binding proteins, and proteins resembling but not identical with troponin C have been isolated from several nonmuscle cells (45–49a). Troponins T and I have not, however, been purified from these cells.

α-Actinin

The actin fibers of the striated-muscle sarcomere insert into a plate of protein, the "Z-line" (Figure 4). α-Actinin, a 200,000 dalton protein that binds to F-actin and is composed of noncovalently joined subunits of identical size, comprises

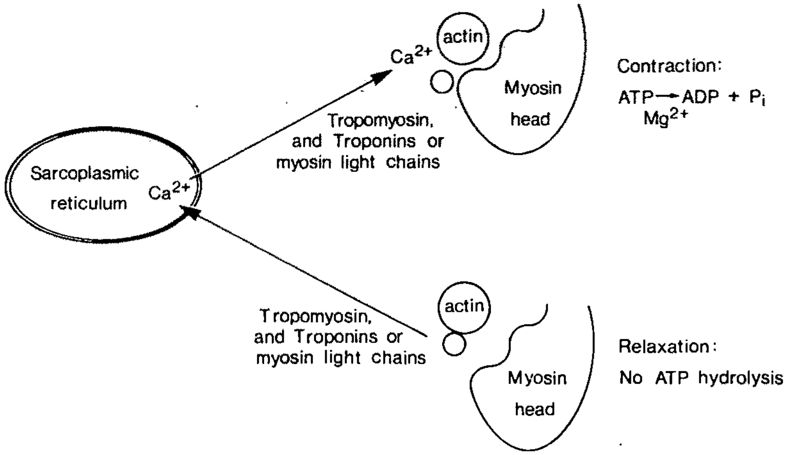


Figure 3 Schematic conception of the control of actin-myosin interaction by calcium and regulatory proteins in striated muscle.

the Z-line. α-Actinin has so far only been isolated from skeletal and smooth muscle (50, 51), but antibodies raised against striated-muscle α-actinin have reacted with various zones of nonmuscle cells (52, 53).

β-Actinin

A minor component of striated muscle called β-actinin inhibits the lengthening of actin polymers by the addition of G-actin monomers (54). No proven function has emerged for this protein in muscle, but a very similar protein isolated from the slime mold *P. polycephalum* (55) is responsible for some peculiar properties of the actin from that organism (56). β-Actinin is a 90,000 molecular weight protein composed of identical 45,000-molecular-weight subunits. When associated with actin filaments, β-actinin causes them to be unusually flexible. This flexibility arises from interruptions in the continuity of the actin polymer that occur when the actin polymerizes in the presence of β-actinin and magnesium chloride (57). Therefore, this altered actin was originally named magnesium polymer (56).

Actin-Binding Protein and Filamin

A high-molecular-weight actin-binding protein was originally purified from rabbit lung macrophages (58) and subsequently from human leukemic granulocytes (59). A structurally similar protein called filamin was described in vertebrate smooth muscle (60, 61). Actin-binding protein and filamin have two 270,000-molecular-weight subunits (58, 65, 76). The only function ascribed to these proteins is their capacity to cross-link actin filaments (62) and form a gel of F-actin (63, 64). The gelling of actin by actin-binding proteins occurs at low actin-

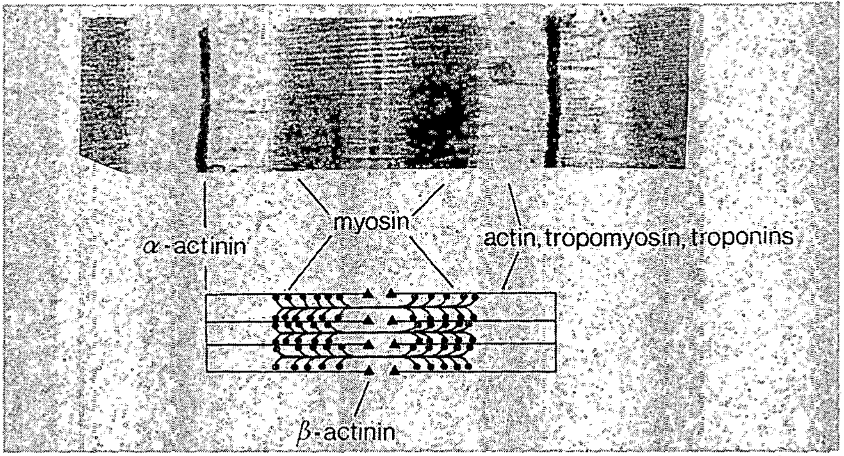


Figure 4 Microscopic appearance and schematic conception of the organization of a striated-muscle sarcomere. The locations of α- and β-actinins, troponins, and tropomyosin have been determined by means of selective extractions and by staining with fluorochrome-labelled specific antibodies (230).

binding protein concentration (1 *M* of actin-binding protein per >680 *M* of actin) and is inhibited by low temperatures (63, 65) and by cytochalasin B (66). Filamin also cross-links F-actin, but much higher concentrations are required to gel actin (64, 67). Cytochalasin B has no effect on this interaction (64). Antibodies raised against smooth-muscle filamin react with a variety of nonmuscle cells (60), and polypeptides, approximately the size of the actin-binding protein subunit, are associated with actins from sea urchin eggs (68), HeLa cells (69), platelets (70), and amoebas (71, 72).

Other Contractile Proteins

Some muscle contractile proteins have not yet yielded definitive counterparts in nonmuscle cells. One is C-protein, a 140,000-molecular-weight polypeptide that influences the packing of muscle myosin filaments (73). Another myosin-associated protein, paramyosin, is involved in a "catch" mechanism by which these animals sustain prolonged isometric contractions (74).

Conversely, some contractile proteins of nonmuscle cells have no defined functions in striated muscle.

SPECTRIN Spectrin is the major protein of erythrocyte membranes (75). It bears certain structural similarities to myosin and to actin-binding proteins, but is clearly distinct from them (63, 76). Spectrin is a contractile protein only by inference. The best evidence for a contractile role of this protein is its apparent association with actin in erythrocyte membranes (77, 78).

PHOSPHORYLATION OF CONTRACTILE PROTEINS Cells have enzyme systems that can add and subtract phosphorus molecules from protein acceptor sites, usually serine residues. In some cases, the controlled phosphorylation and dephosphorylation of protein constitutes a well-characterized system for regulation of cell functions (79). Myosin light chains can be enzymatically phosphorylated, and these reactions can influence the ATPase activity of the protein from smooth-muscle and some nonmuscle cells (80–83).

COFACTORS, CALCIUM-SENSITIZING FACTORS, AND ACTIN-ASSOCIATED PROTEINS Proteins called cofactors are required for the maximal magnesium ATPase and concomitant force-generating capacity of certain cytoplasmic myosins in the presence of actin (62, 84). These positive control factors are in contrast to the negative control proteins like tropomyosins and the troponins of muscle that inhibit the actin-activated magnesium-ATPase activity of myosins. Inhibitors of the actomyosin magnesium-ATPase also exist in some nonmuscle cells. As in vertebrate striated muscle, calcium reportedly releases the inhibition imposed by these factors, hence they are considered calcium-sensitizing factors (39, 85–89). A low-molecular-weight polypeptide identified in association with actin from spleen cells reversibly inhibits the polymerization of the actin (90). Other polypeptides may control the state of actin in cells (91), but these factors are insufficiently characterized to evaluate.

CONTRACTILE PROTEINS IN MUSCLE CELL FUNCTIONS

Striated-Muscle Movement

The controlled shortening of the striated-muscle (skeletal and cardiac) sarcomere is the end result of a remarkable, cooperative interaction between actin and myosin filaments (92). Thin actin filaments attached to α-actinin at the Z-line face each other in opposing arrays parallel to the long axis of the sarcomere, across an actin-free zone. This zone is a parallel assembly of bipolar myosin filaments (Figure 4). In resting or relaxed muscle, only a small area of overlap exists between the two types of filaments. In the overlap zones, the filaments surround each other in a hexagonal herringbone pattern such that each actin filament surrounds a myosin filament, and vice versa. When muscle contracts, the head groups of myosin molecules extending out from the ends of the thick filaments bind to the actin filaments and move them inward by a process energized by splitting ATP. Actin is not merely passive in this process but, upon binding to myosin, activates the the magnesium-ATPase activity of myosin, which transduces the energy into force for movement. The detailed mechanism of this interaction is complex. It involves a cycle wherein a bond forms between the myosin head and actin monomers as the myosin head splits ATP to ADP and phosphorus. The myosin head extends from its anchorage in the filament by a segment of the myosin rod that is flexible enough to permit movement. The head bends toward the center of the sarcomere as the products of ATP hydrolysis dissociate from the complex. A new ATP molecule binds to myosin and dissociates it from actin, which permits a new cycle to begin as the head regains its original position (93). The polarity of actin filaments is such that myosin can only move them away from the Z-line.

Relaxation is the cessation of actin contraction, which permits the sarcomere to lengthen passively. For contraction to cease, actin must stop making contact with the myosin heads. The mechanism of breaking this contact varies: In vertebrate striated, and many other muscles, the tropomyosin-troponins complex is responsible (41–43). Since tropomyosin and the troponins adorn the thin actin filaments, this form of regulation is actin-linked. Only when calcium binds to troponin C does tropomyosin move into a position to permit myosin-actin contact, and only then is the myosin–magnesium ATPase activable by actin. Certain invertebrate muscles have tropomyosin but lack troponins. In these muscles one of the myosin light chains binds calcium and controls myosin contact and activation by actin (94). This is called myosin-linked regulation. Interestingly, certain light chains of myosins usually controlled by actin-linked systems can in hybridization experiments of myosins from myosin-linked control systems, substitute for "calcium" light chains. Finally, some insect and arthropod muscles appear to have both actin- and myosin-linked control systems (94).

Whatever the linkage of control proteins to actin, to myosin, or to both, the ultimate regulator of striated-muscle contraction is ionized calcium. In striated muscle as in most cells, the normal concentration of ionized calcium is very low, relative to magnesium ions. A convoluted cisternal system, the sarcoplasmic reticulum, which insinuates itself among the sarcomeres, exists for the release and uptake of calcium, and this system receives its control from neural impulses (95). Cytoplasmic free calcium concentrations in the micromolar range induce striated-muscle contraction. The central and well-established role of ionized calcium in the control of striated-muscle contractile-protein function has strongly influenced (perhaps wrongly) the thinking of investigators probing the control of contractile-protein interactions in nonmuscle cells.

Smooth-Muscle Movement

Despite its importance, there is much less knowledge of the contractile-protein interaction of the unstriated muscles that surround hollow viscera and blood vessels. Actin filaments extend more or less randomly in the long axis of smooth muscle (96). The actin filaments do not insert into a unit like the Z-line of striated muscle, but attach to electron-dense structures (dense bodies) associated with the plasma membrane and with a network of the intermediate 10-nm filaments (97, 98). α-Actinin has been purified from smooth muscle (51), and the dense bodies react with fluorescein-labelled antibody against striated-muscle α-actinin, which gives them a possible functional connection with the striated-muscle Z-line (99). At least some of the actin exists in a latticework, but the parallel order of actin in striated muscles is lacking, and there is no consistent lattice order throughout a given smooth-muscle cell (96).

Whereas the actin content of smooth muscle is higher per cell than that of skeletal muscle (100), the myosin concentration of smooth muscle is about one fifth that of striated muscle (100). Myosin of smooth muscle forms bipolar filaments when purified in vitro (101, 102), and myosin filaments are identifiable in smooth-muscle cells (96 97), although the orientation of these filaments is a matter of some debate. The consensus seems to be that contractile units of random order exist in smooth muscles, and actin and myosin interact in these units. The mechanism for generation of force is similar to that existing in striated muscle (96, 97, 103).

Supporting the case for at least remotely comparable contraction mechanisms between smooth and striated muscle is the fact that electrical impulses initiate both types of muscle movement, and both muscle cell types contain a sarcoplasmic reticulum (104). Confirmation of this similarity requires the demonstration of a calcium control for smooth-muscle actin-myosin interaction. This demonstration has not been simple, and suggestions for alternative controls have appeared. Regulation of the smooth-muscle actomyosin ATPase by calcium was reported, although the actual actin-activated myosin–magnesium-ATPase activities were relatively low. The calcium regulation was linked to the myosin itself, which is consistent with the failure to identify troponins in smooth muscle (105, 106). Tropomyosin is present in smooth muscle and exists in about the same concentration, with respect to actin, as found in striated muscle (100).

The evidence so far supports a myosin-linked, calcium-regulated mechanism for the control of actin-myosin interaction in smooth muscle, similar to that existing in molluscan muscles (94, 105, 106). The low magnesium-ATPase activity of smooth muscle in the presence of actin has plagued progress in this area. The possible reasons for the slow activity include: (*a*) Smooth-muscle myosin is intrinsically less active enzymatically (107). (*b*) Proteolysis or other detriments hurt the myosin during purification. (*c*) Other factors may be needed to maximize activity. Some preliminary accounts suggest that phosphorylation of smooth-muscle myosins occurs, and that it enhances the magnesium-ATPase activity in the presence of actin and its sensitivity to calcium control (81, 82, 108).

Only guesses exist on how muscle cells actually contract (109). Furthermore, the role of various other contractile proteins such as filamin, which constitutes 6% of chicken gizzard smooth muscle (61, 76), still require integration into the contractile mechanism.

CONTRACTILE PROTEINS IN NONMUSCLE CELL STRUCTURE AND FUNCTION

General Considerations

Striated-muscle cells have the very specialized function of generating linear force, and the arrangement of contractile proteins in them subserves this function. Smooth muscles have a similar purpose, although the force is not linear, and the arrangement of contractile proteins is unclear. In both cases, however, the function of contractile proteins is contractile in the sense of "something becoming smaller"—either the distance between Z-lines or the ends of a smooth-muscle cell. Nonmuscle cells are also specialized for certain functions, and in many instances contractile proteins may regulate these functions. The action of contractile proteins, however, may not always be to contract a cell or even part of it, but rather to provide structural rigidity to a cell or a part thereof, a barrier to the free movement of organelles, and perhaps only occasionally to mediate the motility of the cell or a portion of it.

The Cortical Plasmagel

Many cells viewed in the light microscope have a peripheral layer of organelle-excluding cytoplasm termed hyaline ectoplasm by early observers [the contractile plasmagel of W. H. Lewis (1)]. This ectoplasm comprises surface specializations such as villi, pseudopodia, veils, long processes, and other protuberances important for cell functions. It also appears to limit the access of internal organelles from the plasmalemma. This last feature is evidence for a gel state of the ectoplasm, since it suggests the presence of a rigid structure that prevents the entrance of these organelles. In some cells, organelles move freely over long distances in the internal cytoplasm (endoplasm), which indicates that the failure of organelles to enter the ectoplasm is not because of restraints on the movement of the organelles but because of an ectoplasmic property (110).

Further evidence for the gel state of ectoplasm arises from experiments in which whole cells subjected to centrifugal forces exhibit internal particles that rearrange but still fail to enter the ectoplasm (111) until sufficient shear is applied to break the cell (112).

The ectoplasmic layer seen in the electron microscope is a meshwork of rather randomly oriented thin filaments that bind heavy meromyosin and are therefore actin (3). A three-dimensional fibrous network with cross-linkage defines a gel (113). F-actin purified from striated muscle in solution behaves as if it had some weak network structure.[3] This indicates that weak forces bind F-actin strands at crossover points, but very minimal shear disrupts these contacts (114). Actins purified from most nonmuscle cells behave very similarly, not surprising in light of the similarity of primary and secondary structures (28, 115). Crude cytoplasmic extracts from a broad range of cell types will, when warmed, form gels composed of actin strands (59, 63, 68, 69, 71, 72). These crude gels contain numerous proteins besides actin. Any agent that can cross-link actin filaments can theoretically form an actin gel. Accordingly, myosin (116), α-actinin (117), actin-binding protein (63), four low-molecular-weight gelation factors isolated from an amoeba (67), and undefined contaminants (71, 118) make pure muscle F-actin gel, and all of these proteins could be present and active in cytoplasm. Of these proteins, however, the effect of actin-binding protein resembles the gelling behavior of cytoplasmic extracts most accurately in requiring temperature elevation (63). Furthermore, the concentration of actin-binding protein and of amoeba gelation factors required to gel actin are much lower than those of the other proteins. Purification of actin-binding proteins in the presence of divalent cations causes it to break down into low-molecular-weight fragments that retain the capacity to gel actin (65). This suggests that the amoeba gelation factors may be fragments of actin-binding proteins.

Cytochalasin B, at concentrations lower than or equal to those that influence cell shape, motility, and other functions, dissolves crude cytoplasmic gels (66, 119) as well as gels prepared from purified actin and actin-binding protein (66). These concentrations of cytochalasin B have relatively little or no effect on the structures of actin or myosin or on the interaction of these two proteins (7, 66, 120). Cytochalasin B, therefore, appears to be a reagent that specifically probes the relevance of the gel state of actin cells, rather than the global function of contractile proteins. This conclusion is consistent with cytologic findings in many experimental systems (7), some of which are summarized below.

Internal Plasmasol Gel-Sol Transformations and Other Phenomena

The internal cytoplasm appears in some cells to be less rigid than the cell cortex. Although internal fluidity and impressive endoplasmic streaming occur in some of these cells, in others the endoplasm is apparently at least as rigid as the ectoplasm. The physiologic relevance, if any, for the venerable notion that gel-sol transformation plays some part in cell functions (121) may be restricted to certain cells only. A sol-to-gel change could result from the cross-linking of preexisting actin fibers, or from the simultaneous assembly of fibers from monomers and the cross-linking of these fibers as they polymerize. The former alternative is suggested by old and new experiments that indicate the plasmasol is not a true Newtonian fluid but rather is structured and elastic (122). Actin filaments probably impart much, if not most, of the structure to this material (123). However, some investigators found large amounts of actin in extracts of nonmuscle cells that did not polymerize under conditions that usually polymerized purified actins (124–126). Also, under certain conditions, actin in cytoplasmic extracts can be made to polymerize and gel reversibly and simultaneously (63, 71), although these conditions are not physiological. One possible interpretation of these findings is that actin exists in the cell as a monomer awaiting some functional signal to polymerize. Support for the idea that actin may polymerize abruptly to produce changes in cytoplasmic shape, as in the rapid extension of processes from activated echinoderm sperm, arises from strategically obtained electron micrographs. These show bundles of parallel actin filaments in the newly formed appendages, whereas such filaments did not previously exist. Although these micrographs, embellished with beguiling narrative and simple chemistry, are impressive, the appealing concept that actin polymerization controls cell shape remains to be proven (14, 90, 91, 127, 128).

---

[3] Long fibrous polymers like actin are viscous but do not necessarily form gels. The viscosity arises from initial resistance of randomly oriented fibers to flow, even if there is no interaction between fibers. If the fibers are linked, however, flow can only occur when the linkages break.

Cytoplasmic Movement

Important evidence that contractile proteins control cell movement is summarized as follows:

1. Contractile proteins are probably in all eukaryotic cells.
2. Purified cytoplasmic actins function nearly identically to muscle actins. They share the capacity to activate the magnesium ATPase of muscle myosin at physiologic conditions of pH and salt concentration, an activity relevant for muscle movement.
3. Many cytoplasmic myosins have, like muscle myosins, a low magnesium-ATPase activity, which actin can increase under appropriate conditions.
4. Mixtures of purified cytoplasmic actins, myosins, and relevant control factors undergo contractions in the presence of magnesium and ATP.
5. Cytochalasin B (which dissolves cytoplasmic actin gels) inhibits many cell movements.
6. Contractile proteins, insofar as they are identifiable, are somtimes visible in micrographs in regions of cells undergoing movement.

The details of some of the evidence cited above are discussed more fully below, but the following facts must be emphasized: (*a*) The basic mechanical-chemical interaction between actin and myosin may be similar in muscle and nonmuscle cells, but the very high actin-to-myosin ratios of nonmuscle cells relative to muscle cells indicate that the overall contractile mechanism must be different. The low myosin content puts a severe upper limit on the number of myosin filaments that can exist in a cell (25). Myosin could conceivably work as a molecule without forming filaments. There is evidence that "tailess" myosin can produce movement (71, 129). A gel state of actin may permit contraction of actin filaments by myosin to be transmitted over a greater volume than if myosin were only to react with single actin filaments (63). Finally, contraction might only occur in specific regions of the cell where myosin is concentrated. (*b*) The mechanics of contractile-protein interaction are not definitely known for any cell type other than striated muscle. (*c*) The control of actin-myosin interaction is not worked out in any cell type except muscle, where calcium regulates by means of control proteins.

The criteria for establishing calcium control of nonmuscle cell contractile proteins include: (*a*) an influence of calcium or calcium ionophores on the functions of intact cells; (*b*) some sort of calcium-sequestering system analogous to sarcoplasmic reticulum, or else functionally related release or transport of calcium from the other membranes to the contractile apparatus; (*c*) an influence of calcium on the magnesium-ATPase activity or movement of cells made permeable with glycerol or other agents (130), of cytoplasmic extracts (71, 131), cell fragments (132, 133), or crude actomyosin preparations (134); (*d*) control of purified actomyosin–magnesium-ATPase activity and/or movement by calcium in the presence of relaxing factors derived from the cell in question (85–89); and (*e*) reconstitution of a calcium-activated actomyosin with purified relaxing proteins. Of these criteria, the last is the *sine qua non*, and no cytoplasmic contractile system has yet fulfilled it. However, preliminary evidence already obtained for calcium control of cytoplasmic contractile proteins should encourage further search.

Location of Contractile Proteins

Current histological techniques involving both light and electron microscopy provide only the vaguest guesses as to the location of contractile proteins in cells. Such techniques can localize the proteins when they are aggregated into bundles, but they do not reassure us that important molecules have not escaped detection elsewhere. Chemical methods are more sensitive but less specific. F-actin is a large and abundant fibrous protein with many associated molecules. Small amounts of this complex can easily contaminate subcellular fractions undergoing analysis. Even with these reservations, the evidence is strong that contractile proteins have some noncovalent association with surface membranes (28, 135–141). Indeed, logic dictates that, if contractile proteins are to move anything, they must make contact with this organelle to pull or push it. The erythrocyte is an advantageous model, because its membrane is easy to isolate. All of the erythrocyte actin and the actin-associated protein, spectrin, reside beneath the plasma membrane. The evidence that these proteins have some function related to the structure of the membrane includes these observations: (*a*) Manipulations that aggregate spectrin, e.g. incorporation of antibodies against spectrin into erythrocytes or else lowering of the intracellular pH, causes changes in erythrocyte membrane shape (142, 143). (*b*) Depletion of erythrocyte membranes of the spectrin-actin complex by extraction in low-ionic-strength buffers causes membrane fragmentation (75, 144). (*c*) The spectrin of ATP-depleted erythrocytes becomes inextractable as the membranes lose their deformability and undergo shape changes (145). (*d*) Cytochalasin B increases the osmotic fragility of erythrocytes (145a).

Contractile proteins are clearly not only associated with the plasma membrane, but can bind to membranes of other subcellular organelles (146–148), and in some instances seem to exist throughout the cytoplasm.

Contractile Proteins in Some Specific Cell Functions

CELL LOCOMOTION Actively motile amoeboid cells, which include slime molds and the leukocytes of higher organisms, are a logical and fruitful place to examine the role of contractile proteins in cell movement. The results strongly support the idea that contractile proteins drive amoeboid locomotion: (*a*) Actin (59, 136, 149–155), myosin (59, 151, 155–160), cofactors (62, 84), β-actinin (55), actin-binding protein (59), and calcium-sensitizing factors (88, 86) have been isolated from various amoebas or from leukocytes. (*b*) Understanding of the mechanism of these interactions is emerging. Actin-binding proteins or other gelation factors cause amoeba and leukocyte actins to gel (56, 57). Amoeba and leukocyte actins activate the magnesium ATPase of muscle myosin (59, 136, 150) and of amoeba and leukocyte myosins (136, 155). In one species examined, a protein cofactor was required for the activation of amoeba myosin–magnesium ATPase by actin (84). In reconstituted mixtures of purified proteins, calcium does not influence the magnesium-ATPase activity (161), but crude preparations of certain amoebas reportedly confer calcium regulation on such a mixture (85, 86). The magnesium-ATPase activity of crude actomyosin preparations of some amoebas and leukocytes can be activated by calcium ions (85, 134). Extracts of amoebas and leukocytes underwent temperature-dependent actin assembly (59, 71, 72, 162), gelation (59, 7, 72), streaming (163), and movements that the observers believed resembled the movements of pseudopods. These movements were sensitive to the calcium concentration (131). (*c*) Cytochalasin B reversibly inhibits the locomotion of leukocytes (164). (*d*) A genetic anomaly of contractile proteins in human polymorphonuclear leukocytes is the only example thus far of a contractile-protein disorder in human disease. Polymorphonuclear leukocytes of a male infant were very defective in spreading, phagocytosis, and especially translational movement. Actin from these cells was defective in assembly relative to normal polymorphonuclear leukocyte actin (165).

Fibroblasts and other cells maintained in tissue culture are less active but nonetheless motile. The phenomenology of their spreading and locomotion is well-described (166). Actin (167) and myosins (168–170) have been purified from these cells, and antibodies to muscle tropomyosin (44), α-actinin (53), and filamin (60) react with them. Cytochalasins inhibit their movement and induce peculiar morphologic changes (7). These cells differ morphologically from amoeboid cells in forming actin filament bundles or "stress fibers" at points of attachment (166, 171), and these stain beautifully with antibodies against actin, myosin, filamin, and tropomyosin (30, 33, 44, 53, 60, 172, 173) (Figure 5). In one type of cultivated cell, the staining pattern of antibody to α-actinin seemed to suggest it served as a binding site for stress fibers forming in cells as they spread (53). In other experiments, cytochalasins caused stress fibers to break into small segments that stained with antibodies against actin (174). In summary, the importance of contractile proteins in the mobility of these cells seems certain.

AXONAL TRANSPORT AND GROWTH Actin (126, 175), myosin (176–178), tropomyosin (37, 39), troponin C–like proteins (45, 46, 48), and relaxing factors (39, 89) have been isolated from brain or other neural tissues. The cells that comprise these organs (with the possible exception of glia) are not known for their mobility. The proteins may rather participate in axonal transport (179, 180) or neural appendage growth (180).

MITOSIS During mitosis, motile events include the movement of chromosomes, the fission of the dividing cell at the contractile ring, and the separation of the daughter cells. The following evidence suggests involvement of contractile proteins in some aspects of cell division: (*a*) the apparent presence of actin fibers and myosin between chromosomes and the mitotic poles in the spindle region (5, 30, 181–183); (*b*) the presence, and possibly concentration, of actin and myosin in the cleavage furrow or contractile ring (30, 184–186); (*c*) the apparent concen-

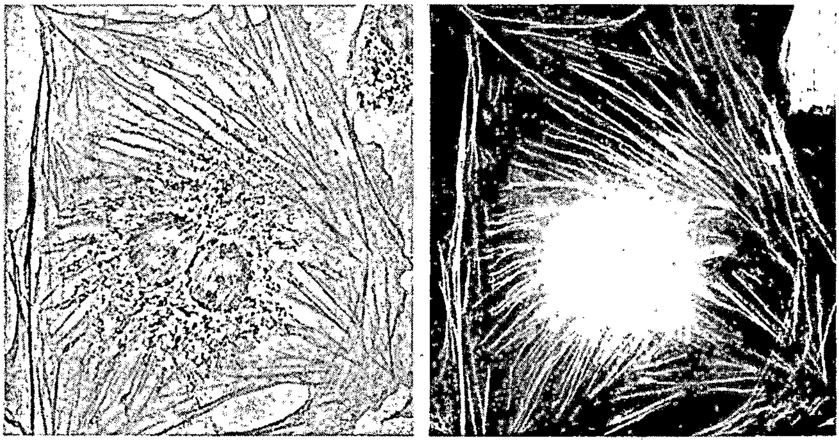


Figure 5 Spread fibroblast showing (*left*) stress fibers seen in the phase contrast microscope and (*right*) stained with fluorescence microscope (reprinted with permission from Ref. 173).

tration of actin in the pseudopodia that develop as the daughter cells separate (186); (*d*) the orientation of actin fibers in dividing cells, which may program the orientation of the daughter cells (187); (*e*) the fact that cytochalasin B dissolves the contractile ring of dividing cells, which is consistent with an actin gel participating in or powering the act of fission (184); (*f*) antibodies against starfish egg myosin inhibit cell division when injected into the egg (187a).

PLATELET FUNCTION Drugs influencing contractile proteins in mitosis could provide pharmacologic control of cell division that occurs in developmental and neoplastic processes in man. Platelets contain actin (124, 188), myosin (25, 189, 190), tropomyosin (36), and actin-binding protein (70). Stimulated platelets undergo a series of motile and contractile events including pseudopod extension, secretion of cytoplasmic granules into the medium, and clot retraction. Prior to stimulation of the platelet, no contractile structure is discernible, but after activation, actin filaments and possibly myosin filaments are easily visible in the cytoplasm, in pseudopodia, and outside the cell as well (128, 191). Cytochalasin B inhibits platelets' pseudopod formation and clot retraction (192), which suggests that the gel state of actin may also be important for these functions. One hypothesis of contractile-protein mediation of platelet function (Figure 6) claims that activation of the platelet causes monomeric actin to polymerize into parallel filaments, thereby driving long processes outward from the cell. The actin-containing processes bind to fibrin. Myosin in the actin cell body causes the actin filament of the processes to slide inward, thereby effecting centrifugal compression of the fibrin meshwork (193). The polarity of actin filaments in platelet pseudopods is such that this directional movement is plausible (193a). The presence of tropomyosin and a troponin C–like protein in platelets (36, 38) raises the possibility that calcium regulates myosin-actin interaction in platelet movement, but only preliminary work with crude protein fractions supports this idea (87). Removal of extracellular calcium inhibits clot retraction by intact platelets (194), but the concentration of extracellular calcium need not be relevant to the control of intracellular contractile events. Cofactor-mediated (193) and phosphorylation-mediated (83) regulation of platelet actomyosin–magnesium-ATPase activity have also been proposed.

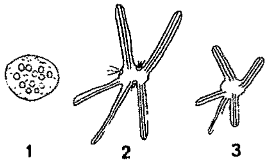


Figure 6 Fantasy about platelet functions possibly mediated by contractile proteins. Platelets when activated extend pseudopodia, secrete granules, and then retract pseudopodia. 1. The cortical actin gel maintains the resting platelet in a discoid shape and keeps its granules from the plasmalemma. 2. Upon activation, focal contraction of the gel by myosin generates pseudopodia and disrupts the gel at other points, permitting granules to approach the plasmalemma and discharge into the medium. 3. Further contraction contracts the pseudopodia [modified from Pollard et al (193)].

ABSORPTION AND SECRETION The discussion of the role of contractile proteins in cell absorption and secretion is divided into three sections; microvillus movement, phagocytosis and pinocytosis, and finally, exocytyosis.

Microvillus Movement The microvilli of intestinal, renal tubular, and other epithelial cells have cords of actin filaments that run parallel to the axis of the microvillus (195–197) (Figure 7). The polarity of these filaments is such that when they bind heavy meromyosin, the arrowheads point away from the tip of the villus (198). Based on the model of striated-muscle contraction, myosin interaction with these filaments would cause them to move away from the villus tips. Antibodies directed against muscle α-actinin react with the villus tips, which suggests that this region is analogous to the Z-line of striated muscle (52). The actin filaments also have cross-bridges that extend to the villus membrane, but the composition of these bridges is uncertain (195, 198). The parallel filaments leave the villus and enter the terminal web, a region where an actin filamentous array is quite random, more like the periphery of unspecialized cells. Electron micrographs of this area reveal structures resembling bipolar myosin filaments, and homogenates of microvilli have a myosin-like ATPase activity that is activated by EDTA and 0.6 *M* KCl. The exact location of these filaments in the terminal web is debated (198, 199). Microvilli of isolated brush borders rendered permeable to the detergent, Triton X-100®, appear to contract or implode into the terminal web upon addition of magnesium and ATP. A role for calcium ions in this reaction is disputed (198, 199). Cytochalasin B does not alter brush border motility or morphology but impedes the development of microvilli in organ cultures (199a, 200).

The structure and function of the microvillus clearly seem to depend on contractile proteins. The relatively structured nature of contractile proteins in this organelle makes it an excellent subject for study.

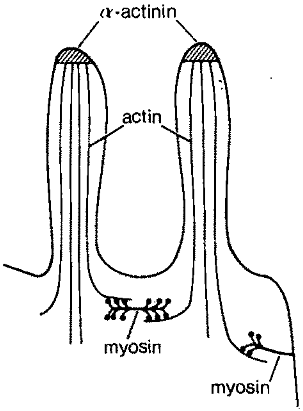


Figure 7 Fantasy of how contractile proteins might control brush-border motility [from Rodewald et al (199) and Mooseker (198)].

Phagocytosis and Pinocytosis The phagocytosis of solid particles and the absorption of fluid by cells are activities that occur in the peripheral plasmagel layer (201). Both processes involve the internalization of membranes as spherical vacuoles or vesicles. The isolation of contractile proteins from actively phagocytic and pinocytic amoeboid cells and leukocytes and from macrophages (58, 62) permits correlation of their interaction during phagocytosis and pinocytosis. During phagocytosis, pseudopodia, which contain principally microfilaments, spread around solid objects (202)(Figure 8). The extractability of actin-binding protein from plasma membranes increases during phagocytosis (63), which may mean that the ingestible particles somehow make actin-binding protein more accessible to actin. Actin-binding protein and myosin appear to advance into the tips of pseudopodia moving around objects (203). During pinocytosis, membrane invaginations appear to be surrounded by filaments. Cytochalasin B inhibits phagocytosis and macropinocytosis, which indicates that an actin gel is necessary for both processes to occur optimally (204, 205).

Exocytosis Secretion of cell products encased in granules to the outside medium is a function of exocrine and endocrine pancreas, salivary gland, mast cells, and other cells, and morphologically is endocytosis in reverse. The plasmagel layer appears to dissolve focally, permitting internal granules to gain access to the plasmalemma (206). The idea of the cortical gel as a barrier to granule egress gains support from data that shows cytochalasin B enhances the exocytosis-mediated secretion of several cell types (207–209). Granule secretion was also excessive in leukocytes with a defective actin polymerization (165).

Contractile proteins have been isolated from cells specializing in secretion by exocytosis. These include typical cytoplasmic myosins from adrenal, pituitary, and pancreas (210, 211), tropomyosin from pancreas (38), and troponin C–like proteins from adrenal glands (47). To the list can be added the contractile proteins discovered in leukocytes, platelets, and even brain, since these cells can undergo exocytosis. Ways in which contractile proteins could activate exocytosis include the overlapping of actin and myosin filaments, as in muscle,

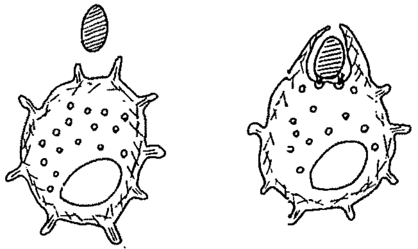


Figure 8 Fantasy of how contractile proteins might mediate phagocytosis by a macrophage. Recognition of an object causes contraction of the cortical actin gel, which forms pseudopodia. The contraction disrupts the gel at the point of original object-cell contact, permitting granule secretion into the developing phagocytic vacuole [from Hartwig et al (203)].

to draw organelles to the plasma membrane (136); or myosin compression and disruption of an actin gel network, which would permit granules to diffuse toward the plasma membrane (203). Some interesting experiments suggested that myosin immobilized on plastic spheres caused synaptosomes to release their contents. A relaxing complex isolated from brain inhibited this release unless free calcium was present. The interpretation of these experiments was that synaptosome actin might interact with plasma-membrane myosin in a manner contolled by free calcium to discharge the synaptosome contents (89).

CANCER The morphology, motility, and "social behavior" of cancer cells clearly differ from most normal cells (212). As suggested elsewhere in this review, all of these attributes of cells are potentially ascribable to the interaction of contractile proteins. Therefore, it is inviting to see if the neoplastic state brings with it definable changes in contractile proteins.

Most work on this question deals with cultured cells transformed with tumor viruses. Although the relevance of specific phenotypic characteristics of transformed cells to those of naturally neoplastic cells can be debated, transformed cells are a reasonable facsimile of natural neoplasia (213). The following findings have emerged: (*a*) Cultured cells transformed by viruses cease to demonstrate the bundles of stress fibers that react with antibodies against actin or myosin (214–220). (*b*) Despite the loss of filament bundles, electron micrographs appeared to reveal more microfilaments in a lattice form in chemically induced or spontaneous skin tumors (221, 222) or in virally transformed cells (218) than in normal or untransformed cells; and various human malignant tumors bound more antibody to smooth-muscle actin than did homologous normal tissues (223, 224). (*c*) The proportion of actin extractable from cells with dilute buffers was greater in transformed than in untransformed cells (225) and, conversely, the quantity of "membrane" actin and myosin was lower in transformed cells (139, 226).

The most conservative interpretation of these observations is that they simply reflect characteristics of very motile cells. Leukocytes tend not to have stress fibers, and their social behavior resembles that of tumor cells (227). Indeed, the loss of actin bundles correlated better with deficiencies of cell adherence to a surface than with the transformed state per se (228). However, the interaction of contractile proteins and tumor cells could differ from that of normal cells, and appropriate pharmacologic strategy might take advantage of such differences. Indeed, cytochalasin B rendered cultured tumor cells more susceptible than normal cells to the toxic effects of cytosine arabinoside, but unfortunately, the drug was not an effective tumoricidal agent in vivo (229).

CONCLUSIONS

1. The cytoplasm of cells is a colloidal network of contractile proteins. Actin filaments are the major components of this network. Other contractile proteins interact with these filaments to create structural rigidity and movement.
2. The structure and function of contractile proteins in striated muscles is well

characterized and thus provides a good example for extrapolation to an analysis of contractile-protein structure and function of nonmuscle cells. However, the interaction of contractile proteins of various cells may be unique.

3. The study of contractile proteins in cells other than muscle has distinct difficulties: (*a*) The proteins are present in much lower concentration than in muscle, and only a few cell types are obtainable for study in quantities comparable to muscle. (*b*) Proteolysis and other detriments may be more severe in nonmuscle cells. (*c*) The organization of contractile proteins is difficult to define in nonmuscle cells. (*d*) The effort is diffuse; investigations examine a wide variety of different, or less commonly the same, cells.

Reflecting these difficulties, a catalogue of putative nonmuscle cell functions related to contractile proteins reveals no one system to be definitively worked out. Nevertheless, the ubiquity of contractile proteins and the obvious importance of their interactions presages increasing relevancy for physiology and medicine.

Literature Cited

1. Lewis, W. H. 1939. The role of a superficial plasmagel layer in changes of form, locomotion, and division of cells in tissue cultures. *Z. Exp. Zellforsch.* 23: 1–7
2. Loewy, A. G. 1952. An actomyosin-like substance from the plasmodium of a myxomycete. *J. Cell. Comp. Physiol.* 40: 127–56
3. Ishikawa, H., Bischoff, R., Holtzer, H. 1969. Formation of arrowhead complexes with heavy meromyosin in a variety of cell types. *J. Cell Biol.* 43: 312–28
4. Huxley, H. E. 1963. Electron microscope studies on the structure of natural and synthetic protein filaments from striated muscle. *J. Mol. Biol.* 3: 281–308
5. Aronson, J. F. 1965. The use of fluorescein-labeled heavy meromyosin for the cytological demonstration of actin. *J. Cell Biol.* 26: 293–98
6. Carter, S. B. 1967. Effects of cytochalasins on mammalian cells. *Nature* 213: 261–64
7. Wessells, N. D., Spooner, B. S., Ash, J. F., Bradley, M. O., Ludeña, M. A., Taylor, E. L., Wrenn, J. T., Yamada, K. M. 1971. Microfilaments in cellular and developmental processes. *Science* 171: 135–43
8. Bray, D. 1972. Cytoplasmic actin: A comparative study. *Cold Spring Harbor Symp. Quant. Biol., Vol. 37*, pp. 567–71
9. Goldman, R., Pollard, T., Rosenblaum, J., eds. 1976. *Cell Motility, Vol. 3* Cold Spring Harbor Conf. Cell Proliferation
10. Puszkin, E., Puszkin, S., Lo, L. W., Tanenbaum, S. W. 1973. Binding of cytochalasin D to platelet and muscle myosin. *J. Biol. Chem.* 248: 7754–61
11. Castle, A. G., Marsh, C. W., Crawford, N. 1976. Evidence against tubulin-actin homology. *FEBS Lett.* 68: 263–70
12. Oosawa, F., Kasai, M. 1971. Actin. In *Subunits in Biological Systems, Part A*, eds. S. N. Timasheff, G. D. Fasman. pp. 261–322. New York: Dekker
13. Cooke, R. 1975. The role of the bound nucleotide in the polymerization of actin. *Biochemistry* 14: 3250–56
14. Isenburg, G., Wohlfarth-Bottermann, K. E. 1976. Transformation of cytoplasmic actin. *Cell Tissue Res.* 173: 495–528
15. Lazarides, E., Lindberg, U. 1974. Actin is the naturally occurring inhibitor of deoxyribonuclease I. *Proc. Natl. Acad. Sci. USA* 71: 4742–46
16. Lindberg, U. 1967. Studies on the complex formation between deoxyribonuclease I and spleen inhibitor II. *Biochemistry* 6: 343–47
17. Hitchcock, S. E., Carlsson, L., Lindberg, U. 1976. Depolymerization of F-actin by deoxyribonuclease I. *Cell* 7: 531–42
18. Arnold, H., Pette, D. 1968. Binding

of glycolytic enzymes to structure proteins of the muscle. *Eur. J. Biochem.* 6:163–71

19. Clarke, F. M., Morton, D. J. 1976. Aldolase binding to actin-containing filaments. Formation of paracrystals. *Biochem. J.* 159:797–98
20. Elzinga, M., Maron, B. J., Adelstein, R. S. 1976. Human heart and platelet actins are products of different genes. *Science* 191:94–95
21. Rubenstein, P. A., Spudich, J. A. 1977. Actin microheterogeneity in chick embryo fibroblasts. *Proc. Natl. Acad. Sci. USA* 74:120–23
22. Storti, R. V., Rich, A. 1976. Chick cytoplasmic actin and muscle actin have different structural genes. *Proc. Natl. Acad. Sci. USA* 73:2346–50
23. Gruenstein, E., Rich, A. 1975. Non-identity of muscle and non-muscle actins. *Biochem. Biophys. Res. Commun.* 64:472–77
24. Lowey, S., Slayter, H. S., Weeds, A. G., Baker, H. 1969. Substructure of the myosin molecule I. Subfragments of myosin by enzymic degradation. *J. Mol. Biol.* 42:1–29

24a. Potter, J. D. 1974. The content of troponin, tropomyosin, actin, and myosin in rabbit skeletal muscle myofibrils. *Arch. Biochem. Biophys.* 162:436–41

25. Niederman, R., Pollard, T. D. 1975. Human platelet myosin II. In vitro assembly and structure of myosin filaments. *J. Cell Biol.* 67:72–92
26. Seidel, J. C. 1969. The effects of monovalent and divalent cations on the ATPase activity of myosin. *Biochim. Biophys. Acta* 189:162–70
27. Eisenberg, E., Moos, C. 1967. The interaction of actin with myosin and heavy meromyosin in solution at low ionic strength. *J. Biol. Chem.* 242:2945–51
28. Weihing, R. R. 1976. Occurrence of microfilaments in non-muscle cells and tissues. In *Biological Handbooks, Cell Biology,* Vol. 1, ed. P. L. Altman, D. D. Katz. pp. 341–56. Bethesda, Md.: FASEB Publ.
29. Kessler, D., Nachmias, V. T., Loewy, A. G. 1976. Actomyosin content of *Physarum* plasmodia and detection of immunological cross-reactions with myosins from related species. *J. Cell Biol.* 69:393–406
30. Fujiwara, K., Pollard, T. D. 1976. Fluorescent antibody localization of myosin in the cytoplasm, cleavage furrow, and mitotic spindle of human cells. *J. Cell Biol.* 71:848–75
31. Gröschel-Stewart, U. 1971. Comparative studies of human smooth and striated muscle myosins. *Biochim. Biophys. Acta* 229:322–34
32. Willingham, M. C., Ostlund, R. E., Pastan, I. 1974. Myosin is a component of the cell surface of cultured cells. *Proc. Natl. Acad. Sci. USA 71: 4144–48*
33. Weber, K., Gröschel-Stewart, U. 1974. Antibody to myosin: the specific visualization of myosin-containing filaments in non-muscle cells. *Proc. Natl. Acad. Sci. USA* 71:4561–64
34. Becker, C. G., Nachman, R. L. 1973. Contractile proteins of endothelial cells, platelets and smooth muscle. *Am. J. Pathol.* 71:1–22
35. Bailey, K. 1948. Tropomyosin: a new asymmetric protein component of the muscle fibril. *Biochem. J.* 43:271–82
36. Cohen, I., Cohen, C. 1972. A tropomyosin-like protein from human platelets. *J. Mol. Biol.* 68:383–87
37. Fine, R. E., Blitz, A. L., Hitchcock, S. E., Kaminer, B. 1973. Tropomyosin in brain and growing neurones. *Nature New Biol.* 245:182–86
38. Fine, R. E., Blitz, A. L. 1975. A chemical comparison of tropomyosins from muscle and non-muscle tissues. *J. Mol. Biol.* 95:447–54
39. Mohendran, C., Berl, S. 1977. Isolation of troponin-like complex from bovine brain cortex. *Proc. Natl. Acad. Sci. USA* 74:2273–77
40. Katz, A. M. 1964. Influence of tropomyosin upon the reactions of actomyosin at low ionic strength. *J. Biol. Chem.* 239:3304–11
41. Ebashi, S. 1972. Calcium ions and muscle contraction. *Nature* 240:217–18
42. Weber, A., Murray, J. M. 1973. Molecular control mechanisms in muscle contraction. *Physiol. Rev.* 53:612–73
43. Szent-Györgyi, A. G. 1975. Calcium regulation of muscle contraction. *Biophys. J.* 15:707–23
44. Lazarides, E. 1975. Tropomyosin antibody: the specific localization of tropomyosin in non-muscle cells. *J. Cell Biol.* 65:549–61
45. Fine, R., Lehman, W., Head, J., Blitz, A. 1975. Troponin-C in brain. *Nature* 258:260–62

46. Kuo, I. C. Y., Coffee, C. J. 1976. Purification and characterization of a troponon-C–like protein from bovine adrenal medulla. *J. Biol. Chem.* 251:1603–9
47. Kuo, I. C. Y., Coffee, C. J. 1976. Bovine adrenal medulla troponin-C. Demonstration of a calcium-dependent conformational change. *J. Biol. Chem.* 251:6315–19
48. Watterson, D. M., Harrelson, W. G. Jr., Keller, P. M., Sharief, F., Vanaman, T. C. 1976. Structural similarities between the Ca^{2+}-dependent regulatory proteins of 3′:5′-cyclic nucleotide phosphodiesterase and actomyosin ATPase. *J. Biol. Chem.* 251:4501–13
49. Dedman, J. R., Potter, J. D., Means, A. R. 1977. Biological cross-reactivity of rat testis phosphodiesterase activator protein and rabbit skeletal muscle troponin-C. *J. Biol. Chem.* 252:2437–40
50. Ebashi, S., Ebashi, F. 1965. Alpha-actinin, a new structural protein from striated muscle I. Preparation and action on actomyosin-ATP interaction. *J. Biochem.* 58:7–12
51. Singh, I., Goll, D. E., Robson, F. M., Stromer, M. H. 1977. N- and C-terminal amino acids of purified alpha-actinin. *Biochim. Biophys. Acta* 491:29–45
52. Schollmeyer, J. V., Goll, D. E., Tilney, L., Mooseker, M., Stromer, M. 1974. Localization of alpha-actinin in non-muscle material. *J. Cell Biol.* 63:304a
53. Lazarides, E., Burridge, K. 1975. Alpha-actinin: immunofluorescent localization of a muscle structural protein in non-muscle cells. *Cell* 6: 289–98
54. Maruyama, K. 1965. Some physicochemical properties of beta-actinin, "actin-factor," isolated from striated muscle. *Biochim. Biophys. Acta* 102:542–48
55. Hatano, S., Owaribe, K. 1976. Actin and actinin from myxomycete plasmodia. See Ref. 9, pp. 499–511
56. Hatano, S., Totsuka, T., Oosawa, F. 1967. Polymerization of plasmodium actin. *Biochim. Biophys. Acta* 140:109–22
57. Kamiya, R., Maruyama, K., Kuroda, M., Kawamura, M., Kikuchi, M. 1972. Mg-polymer of actin formed under the influence of beta-actinin. *Biochim. Biophys. Acta* 256: 120–31
58. Hartwig, J. H., Stossel, T. P. 1975. Isolation and properties of actin, myosin and a new actin-binding protein in rabbit alveolar macrophages. *J. Biol. Chem.* 250:5699–5705
59. Boxer, L. A., Stossel, T. P. 1976. Interactions of actin, myosin and an actin-binding protein of chronic myelogenous leukemia leukocytes. *J. Clin. Invest.* 57:964–76
60. Wang, K., Ash, J. F., Singer, S. J. 1975. Filamin, a new high-molecular-weight protein found in smooth muscle and non-muscle cells. *Proc. Natl. Acad. Sci. USA* 72:4483–86
61. Shizuta, Y., Shizuta, H., Gallo, M., Davies, P., Pastan, I., Lewis, M. 1976. Purification and properties of filamin, an actin-binding protein from chicken gizzard. *J. Biol. Chem.* 251:6562–67
62. Stossel, T. P., Hartwig, J. H. 1975. Interactions between actin, myosin and a new actin-binding protein of rabbit alveolar macrophages. Macrophage myosin Mg^{++} adenosine triphosphatase requires a cofactor for activation by actin. *J. Biol. Chem.* 250:5706–12
63. Stossel, T. P., Hartwig, J. H. 1976. Interactions of actin, myosin and a new actin-binding protein of rabbit pulmonary macrophages II. Role in cytoplasmic movement and phagocytosis. *J. Cell Biol.* 68:602–19
64. Wang, K., Singer, S. J. 1977. Interaction of filamin with F-actin in solution. *Proc. Natl. Acad. Sci. USA* 74:2021–25
65. Hartwig, J. H., Stossel, T. P. 1977. Actin-binding protein as an actin-gelling factor: specificity and mechanism. *J. Cell Biol.* 75:MF273
66. Hartwig, J. H., Stossel, T. P. 1976. Interactions of actin, myosin and an actin-binding protein of rabbit pulmonary macrophages III. Effects of cytochalasin B. *J. Cell Biol.* 71:295–303
67. Maruta, H., Korn, E. D. 1977. Purification from *Acanthamoeba castellanii* of proteins that induce gelation and syneresis of F-actin. *J. Biol. Chem.* 252:399–402
68. Kane, R. E. 1975. Preparation and purification of polymerized actin from sea urchin egg extracts. *J. Cell Biol.* 66:305–15
69. Weihing, R. R. 1977. Effects of myosin and heavy meromyosin on actin-celated gelatin of HeLa cell extracts. *J. Cell Biol.* 75:95–103

70. Lucas, R. C., Gallagher, M., Stracher, A. 1976. Actin and actin-binding protein in platelets. In *Contractile Systems in Non-muscle Tissues*, ed. S. V. Perry, A. Margreth, R. S. Adelstein, pp. 133–139. Amsterdam: Elsevier Biomed. Press
71. Pollard, T. D. 1976. The role of actin in the temperature-dependent gelation and contraction of *Acanthamoeba* extracts. *J. Cell Biol.* 68:579–601
72. Taylor, D. L., Rhodes, J. A., Hammond, S. A. 1976. The contractile basis of amoeboid movement. II. Structure and contractility of motile extracts and plasmalemma-ectoplasm ghosts. *J. Cell Biol.* 70:112–43
73. Offer, G., Moos, C., Starr, R. 1973. A new protein of the thick filaments of vertebrate skeletal myofibrils. Extraction, purification and characterization. *J. Mol. Biol.* 74:653–70
74. Szent-Györgyi, A. G., Cohen, C., Kendrick-Jones, J. 1971. Paramyosin and the filaments of molluscan "catch" muscles II. Native filaments: isolation and characterization. *J. Mol. Biol.* 56:239–58
75. Marchesi, V. T., Steers, E. Jr. 1968. Selective solubilization of a protein component of the red cell membrane. *Science* 158:203–4
76. Wang, K. 1977. Filamin, a new high-molecular-weight protein found in smooth muscle and non-muscle cells. Purification and properties of chicken gizzard filamin. *Biochemistry* 16:1857–65
77. Tilney, L. G., Detmers, P. 1975. Actin in erythrocyte ghosts and its association with spectrin. Evidence for a nonfilamentous form of these two molecules in situ. *J. Cell Biol.* 66:508–20
78. Sheetz, M. P., Painter, R. G., Singer, S. J. 1976. Relationships of the spectrin complex of human erythrocyte membranes to the actomyosins of muscle cells. *Biochemistry* 15:4486–92
79. Greengard, P. 1976. Possible role for cyclic nucleotides and phosphorylated membrane proteins in postsynaptic actions of neurotransmitters. *Nature* 260:101–8
80. Perry, S. V. 1976. Regulation of contraction in muscle. See Ref. 70, pp. 141–51
81. Chacko, S., Conti, M. A., Adelstein, R. S. 1977. Effect of phosphorylation of smooth muscle myosin on actin activation and Ca^{2+} regulation. *Proc. Natl. Acad. Sci. USA* 74:129–33
82. Sobieszek, A., Small, J. V. 1977. Regulation of the actin-myosin interaction in vertebrate smooth muscle: activation via a myosin light-chain kinase and the effect of tropomyosin. *J. Mol. Biol.* 112:559–76
83. Adelstein, R. S., Conti, M. A. 1975. Phosphorylation of platelet myosin increases actin-activated myosin ATPase activity. *Nature* 256:597–98
84. Pollard, T. D., Korn, E. D. 1973. *Acanthamoeba* myosin II. Interaction with actin and a new cofactor protein required for actin activation of Mg^{2+} adenosine triphosphatase activity. *J. Biol. Chem.* 248:4691–97
85. Nachmias, V. T., Asch, A. 1976. A calcium-sensitive preparation from *Physarum polycephalum*. *Biochemistry* 15:4273–78
86. Mockrin, S. C., Spudich, J. A. 1976. Calcium control of actin-activated myosin adenosine triphosphatase from *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA* 73:2321–25
87. Cohen, I., Kaminski, E., DeVries, A. 1973. Actin-linked regulation of the human platelet contractile system. *FEBS Lett.* 34:315–17
88. Tanaka, H., Hatano, S. 1972. Extraction of native tropomyosin-like substances from myxomycete plasmodium and the cross reaction between plasmodium F-actin and muscle native tropomyosin. *Biochim. Biophys. Acta* 257:445–51
89. Puszkin, S., Kochwa, S. 1974. Regulation of neurotransmitter release by a complex of actin with relaxing protein isolated from rat brain synaptosomes. *J. Biol. Chem.* 249:7711–14
90. Carlsson, L., Nyström, L. E., Sundkvist, I., Markey, F., Lindberg, U. 1976. Profilin, a low-molecular-weight protein controlling actin polymerisability. See Ref. 70, pp. 39–49
91. Tilney, L. G. 1976. The polymerization of actin. III. Aggregates of nonfilamentous actin and its associated proteins: a storage form of actin. *J. Cell Biol.* 69:73–89
92. Huxley, H. E. 1969. The mechanism of muscular contraction. *Science* 164:1356–66

93. Taylor, E. W. 1977. Transient phase of adenosine of triphosphate hydrolysis by myosin, heavy meromyosin, and subfragment 1. *Biochemistry* 16:732–39

94. Lehman, W. 1976. Phylogenetic diversity of the proteins regulating muscular contraction. *Int. Rev. Cytol.* 44:55–92

95. Martonosi, A. 1971. The structure and function of sarcoplasmic reticulum membranes. In *Biomembranes*, ed. L. A. Manson, pp. 191–256. New York: Plenum

96. Small, J. V., Squire, J. M. 1972. Structural basis of contraction in vertebrate smooth muscle. *J. Mol. Biol.* 67:117–49

97. Ashton, F. T., Somlyo, A. V., Somlyo, A. P. 1975. The contractile apparatus of vascular smooth muscle: intermediate high voltage stereo electron microscopy. *J. Mol. Biol.* 98:17–29

98. Small, J. V. 1977. Studies on isolated smooth muscle cells: the contractile apparatus. *J. Cell Sci.* 24:327–49

99. Schollmeyer, J. E., Furcht, L. T., Goll, D. E., Robson, R. M., Stromer, M. H. 1976. Localization of contractile proteins in smooth muscle cells and in normal and transformed fibroblasts. See Ref. 9, pp. 361–88

100. Murphy, R. A., Herlihy, J. T., Megerman, J. 1974. Force-generating capacity and contractile protein content of arterial smooth muscle. *J. Gen. Physiol.* 64:691–705

101. Wachsberger, P. R., Pepe, F. A. 1974. Purification of uterine myosin and synthetic filament formation. *J. Mol. Biol.* 88:385–91

102. Kaminer, B., Szonyi, E., Beccher, C. D. 1976. "Hybrid" myosin filaments from smooth and striated muscle. *J. Mol. Biol.* 100:379–86

103. Small, J. V. 1974. Contractile units in vertebrate smooth muscle cells. *Nature* 249:324–27

104. Somlyo, A. P. 1975. Vascular Smooth Muscle. In *Cellular Pharmacology of Excitable Tissues*, ed. T. Narahishi. pp. 360–407. Springfield, Ill.: Thomas

105. Driska, S., Hartshorne, D. J. 1975. The contractile proteins of smooth muscle. Properties and components of a Ca^{2+}-sensitive actomyosin from chicken gizzard. *Arch. Biochem. Biophys.* 167:203–12

106. Sobieszek, A., Small, J. V. 1976. Myosin-linked calcium regulation in vertebrate smooth muscle. *J. Mol. Biol.* 101:75–92

107. Bárány, M., Bárány, K., Gaetjens, E., Bailin, G. 1966. Chicken gizzard myosin. *Arch. Biochem. Biophys.* 113:205–21

108. Aksoy, M. O., Williams, D., Sharkey, E. M., Hartshorne, D. J. 1976. A relationship between Ca^{2+} sensitivity and phosphorylation of gizzard actomyosin. *Biochem. Biophys. Res. Commun.* 69:35–41

109. Fay, F. S., Delise, C. M. 1973. Contraction of isolated smooth muscle cells—structural changes. *Proc. Natl. Acad. Sci. USA* 70:641–45

110. Kominick, H., Stockem, W., Wohlfarth-Bottermann, K. E. 1973. Cell motility: mechanisms in protoplasmic streaming and amoeboid movement. *Int. Rev. Cytol.* 34:169–249

111. Heilbrunn, L. V. 1956. *The Dynamics of Living Protoplasm.* New York: Academic

112. Angerer, C. A. 1936. The effects of mechanical agitation on the relative viscosity of amoeba protoplasm. *J. Cell. Comp. Physiol.* 8:329–45

113. Ferry, J. D. 1948. Protein gels. *Adv. Protein Chem.* 4:2–78

114. Maruyama, K. 1965. A new protein-factor hindering network formation of F-actin in solution. *Biochim. Biophys. Acta* 94:208–25

115. Gordon, D. J., Yang, Y. Z., Korn, E. D. 1976. Polymerization of *Acanthamoeba* actin. Kinetics, thermodynamics, and co-polymerization with muscle actin. *J. Biol. Chem.* 251:7474–79

116. Szent-Györgyi, A. 1947. *Chemistry of Muscular Contraction.* New York: Academic

117. Maruyama, K., Hama, H., Ishikawa, Y. 1965. Some physicochemical properties of KI-extracted F-actin. *Biochim. Biophys. Acta* 94:200–7

118. Spudich, J. A., Cooke, R. 1975. Supramolecular forms of actin from amoebas of *Dictyostelium discoidium. J. Biol. Chem.* 250:7485–91

119. Weihing, R. R. 1976. Cytochalasin B inhibits actin-related gelation of HeLa cell extracts. *J. Cell Biol.* 71:303–7

120. Spudich, J. A. 1972. Effects of cytochalasin B on actin filaments. See Ref. 8, pp. 585–93

121. Mast, S. O. 1926. Structure, movement, locomotion and stimulation in amoeba. *J. Morphol.* 41:347–425
122. Allen, R. D., Roslansky, J. D. 1959. The consistency of amoeba cytoplasm and its bearing on the mechanism of amoeboid movement I. An analysis of endoplasmic velocity profiles of *Chaos chaos* (L). *J. Biophys. Cytochem.* 6:437–46
123. Pollard, T. D., Korn, E. D. 1971. Filaments of *Amoeba proteus* II. Binding of heavy meromyosin by thin filaments in motile cytoplasmic extracts. *J. Cell Biol.* 48:216–19
124. Probst, E., Lüscher, F. 1972. Studies on thrombosthenin A, the actin-like moiety of the contractile protein from blood platelets. I. Isolation, characterization and evidence for two forms of thrombosthenin A. *Biochim. Biophys. Acta* 278:577–84
125. Abramowitz, J. W., Stracher, A., Detwiler, T. C. 1975. A second form of actin: platelet microfilaments depolymerized by ATP and divalent cations. *Arch. Biochem. Biophys.* 167:230–37
126. Bray, D., Thomas, C. 1976. Unpolymerized actin in fibroblasts and brain. *J. Mol. Biol.* 105:527–44
127. Tilney, L. G., Hatano, S., Ishikawa, H., Mooseker, M. S. 1973. The polymerization of actin: its role in the generation of the acrosomal process of certain echinoderm sperm. *J. Cell Biol.* 59:109–26
128. Zucker-Franklin, D., Grusky, G. 1972. The actin and myosin filaments of human and bovine blood platelets. *J. Clin. Invest.* 51:419–30
129. Oplatka, A., Gadasi, H., Borejdo, J. 1974. The contraction of "ghost" myofibrils and glycerinated muscle fibers irrigated with heavy meromyosin subfragment-1. *Biochem. Biophys. Res. Commun.* 58:905–12
130. Hoffman-Berling, H. 1964. Relaxation of fibroblast cells. In *Primitive Motile Systems in Cell Biology*, ed. R. D. Allen, N. Kamiya. pp. 365–74. New York: Academic
131. Taylor, D. L., Condeelis, J. S., Moore, P. S., Allen, R. D. 1973. The contractile basis of amoeboid movement. I. The chemical control of motility in isolated cytoplasm. *J. Cell Biol.* 59:378–94
132. Hatano, S. 1970. Specific effect of Ca^{2+} on movement of plasmodial fragment obtained by caffeine treatment. *Exp. Cell Res.* 61:199–203
133. Matthews, L. M. 1976. Ca^{++}-regulation in caffeine-derived microplasmodia of *Physarum polycephalum. J. Cell Biol.* 72:502–5
134. Shibata, H., Tatsumi, N., Tanaka, K., Okamura, Y., Senda, N. 1972. A contractile protein possessing Ca^{2+}-sensitivity (natural actomyosin) from leukocytes. Its extraction and some of its propertis. *Biochim. Biophys. Acta* 256:565–76
135. Pollard, T. D., Korn, E. D. 1973. Electron microscopic identification of actin associated with isolated amoeba plasma membrane. *J. Biol. Chem.* 248:448–50
136. Spudich, J. A. 1974. Biochemical and structural studies of actomyosin-like proteins from non-muscle cells II. Purification, properties, and membrane association of actin from amoebae of *Dictyostelium discoideum. J. Biol. Chem.* 249:6013–20
137. Barber, B. H., Crumpton, M. J. 1976. Actin associated with purified lymphocyte plasma membrane. *FEBS Lett.* 66:215–20
138. Gruenstein, E., Rich, A., Weihing, R. A. 1975. Actin associated with membrane from 3T3 fibroblast and HeLa cells. *J. Cell Biol.* 64:223–34
139. Boxer, L. A., Richardson, S., Floyd, A. 1976. Identification of actin-binding protein in membrane of polymorphonuclear leukocytes. *Nature* 263:259–61
140. Gwynn, I. A., Kemp, R. B., Jones, B. M., Gröschel-Stewart, U. 1974. Ultrastructural evidence for myosin of the smooth muscle type at the surface of trypsin-dissociated embryonic chick cells. *J. Cell Sci.* 15:279–89
141. Korn, E. D., Wright, P. L. 1973. Macromolecular composition of an amoeba plasma membrane. *J. Biol. Chem.* 248:439–47
142. Nicolson, G. L., Painter, R. G. 1973. Anionic sites of human erythrocyte membranes II. Antispectrin-induced transmembrane aggregation of the binding sites for positively charged colloidal particles. *J. Cell Biol.* 59:395–406
143. Elgsaeter, A., Shotton, D. M., Branton, D. 1976. Intramembrane particle aggregation in erythrocyte ghosts II. The influence of spectrin aggregation. *Biochim. Biophys. Acta* 426:101–22
144. Lutz, H. U., Lie, S. C., Palek, J.

1977. Release of spectrin-free vesicles from human erythrocytes during ATP depletion I. Characterization of spectrin-free vesicles. *J. Cell Biol.* 73:548–60

145. Lux, S. E., John, K. M., Ukena, T. E. 1977. Diminished spectrin extraction from ATP-depleted human erythrocytes. Evidence that spectrin contributes to erythrocyte deformability. *J. Clin. Invest.* In press

145a. Beck, J. S., Jay, A. W. L., Saari, J. T. 1972. Effects of cytochalasin B on osmotic fragility and deformability of human erythrocytes. *Can. J. Physiol. Pharmacol.* 50:684–88

146. Burridge, K., Phillips, J. H. 1975. Association of actin and myosin with secretory granule membranes. *Nature* 254:526–29

147. Blitz, A. L., Fine, R. E. 1974. Muscle-like contractile proteins and tubulin in synaptosomes. *Proc. Natl. Acad. Sci. USA* 71:4472–76

148. Ostlund, R. E., Leung, J. T., Kipnis, D. M. 1977. Muscle actin filaments bind pituitary secretory granules in vitro. *J. Cell Biol.* 73: 78–87

149. Hatano, S. Oosawa, F. 1966. Isolation and characterization of plasmodium actin. *Biochim. Biophys. Acta* 127:488–98

150. Weihing, R. R., Korn, E. D. 1971. *Acanthamoeba* actin. Isolation and properties. *Biochemistry* 10:590–601

151. Adelman, M. R., Taylor, E. W. 1969. Isolation of an actomyosin-like protein from slime mold plasmodium and the separation of the complex into actin- and myosin-like fractions. *Biochemistry* 12:4964–75

152. Wooley, D. E. 1972. An actin-like protein from amoeba of *Dictyostelium* discoideum. Arch. Biochem. Biophys. 150:519–30

153. Jockusch, B. M., Brown, D. F., Rusch, H. P. 1971. Synthesis and some properties of an actin-like nuclear protein in the slime mold *Physarum polycephalum. J. Bacteriol.* 108:705–14

154. Gordon, D. J., Eisenberg, E., Korn, E. D. 1976. Characterization of cytoplasmic actin isolated from *Acanthamoeba castellanii* by a new method. *J. Biol. Chem.* 251:4778–86

155. Tatsumi, N., Shibata, N., Okamura, Y., Takeuchi, K., Senda, N. 1973. Actin and myosin A from leucocytes. *Biochim. Biophys. Acta* 305:433–44

156. Pollard, T. D., Korn, E. D. 1973. *Acanthamoeba* myosin I. Isolation from *Acanthamoeba castellanii* of an enzyme similar to muscle myosin. *J. Biol. Chem.* 4682–90

157. Hatano, S., Tazawa, M. 1968. Isolation, purification and characterization of myosin B from myxomycete plasmodium. *Biochim. Biophys. Acta* 154:507–19

158. Nachmias, V. T. 1974. Properties of *Physarum* myosin purified by a potassium iodide procedure. *J. Cell Biol.* 62:54–65

159. Stossel, T. P., Pollard, T. D. 1973. Myosin in polymorphonuclear leukocytes. *J. Biol. Chem.* 248:8288–94

160. Clarke, M., Spudich, J. A. 1974. Biochemical and structural studies of actomyosin-like proteins from non-muscle cells. Isolation and characterization of myosin from amoebae of *Dictyostelium discoideum. J. Mol. Biol.* 86:209–22

161. Pollard, T. D., Eisenberg, E., Korn, E. D., Kielley, W. W. 1974. Inhibition of Mg^{++} ATPase activity of actin-activated *Acanthamoeba* myosin by muscle troponin-tropomyosin: implications for the mechanism of control of amoeba motility and muscle contraction. *Biochem. Biophys. Res. Commun.* 51:693–98

162. Pollard, T. D., Ito, S. 1970. Protoplasmic filaments of *Amoeba proteus*. The role of filaments in consistency changes and movement. *J. Cell Biol.* 46:267–89

163. Allen, R. D., Cooledge, J. W., Hall, P. J. 1960. Streaming in cytoplasm dissociated from the giant amoeba, *Chaos chaos. Nature* 187:896–99

164. Zigmond, S. H., Hirsch, J. G. 1972. Effects of cytochalasin B on polymorphonuclear leucocyte locomotion, phagocytosis and glycolysis. *Exp. Cell Res.* 73:383–93

165. Boxer, L. A., Hedley-Whyte, E. T., Stossel, T. P. 1974. Neutrophil actin dysfunction and abnormal neutrophil behavior. *N. Engl. J. Med.* 291:1093–99

166. Harris, A. K. 1973. Cell surface movements related to cell locomotion. In *Locomotion of Tissue Cells*. Ciba Foundation on Symposium 14 (New Series), pp. 3–20. Amsterdam: Elsevier

167. Yand, Y. Z., Perdue, J. F. 1972. Contractile proteins of cultured cells

I. The isolation and characterization of an actin-like protein from cultured chick embryo fibroblasts. *J. Biol. Chem.* 247:4503–9

168. Adelstein, R. S., Conti, M. A., Johnson, G. S., Pastan, I., Pollard, T. D. 1972. Isolation and characterization of myosin from cloned mouse fibroblasts. *Proc. Natl. Acad. Sci. USA* 69:3693–97
169. Ostlund, R. E., Pastan, I., Adelstein, R. S. 1974. Myosin in cultured fibroblasts. *J. Biol. Chem.* 249:3903–7
170. Moore, A., Jaffe, E. A., Becker, C. G., Nachman, R. L. 1977. Myosin in cultured human endothelial cells. *Br. J. Haematol.* 35:71–79
171. Wolosewick, J. J., Porter, K. R. 1976. Stereo high-voltage electron microscopy of whole cells of the human diploid line, W1-38. *Am. J. Anat.* 147:303–24
172. Lazarides, E., Weber, K. 1974. Actin antibody: the specific visualization of actin filaments in non-muscle cells. *Proc. Natl. Acad. Sci. USA* 71:2268–72
173. Goldman, R. D., Lazarides, E., Pollack, R., Weber, K. 1975. The distribution of actin in non-muscle cells. The use of actin antibody in the localization of actin within the microfilament bundles of mouse 3T3 cells. *Exp. Cell Res.* 90:333–44
174. Rathke, P. C., Seib, E., Weber, K., Osborn, M., Franke, W. W. 1977. Rod-like elements from actin-containing microfilament bundles observed in cultured cells after treatment with cytochalasin A (CA) *Exp. Cell Res.* 105:253–62
175. Puszkin, S., Berl, S. 1972. Actomyosin-like protein from brain. Separation and characterization of the actin-like component. *Biochim. Biophys. Acta* 256:695–709
176. Burridge, K., Bray, D. 1975. Purification and structural analysis of myosins from brain and other non-muscle tissues. *J. Mol. Biol.* 99:1–14
177. Ash, J. F. 1975. Purification and characterization of myosin from the clonal rat glial cell strain C-6. *J. Biol. Chem.* 250:3560–66
178. See, Y. P., Metuzals, J. 1976. Purification and characterization of squid brain myosin. *J. Biol. Chem.* 251:7682–89
179. Hoffman, P. N., Lasek, R. J. 1975. The slow component of axonal transport. Identification of major structural polypeptides of the axon and their generality among mammalian neurons. *J. Cell Biol.* 66: 351–66
180. Bray, D. 1973. Model for membrane movements in the neural growth cone. *Nature* 244:354–57
181. Forer, A., Behnke, O. 1972. An actin-like component in spermatocytes of a crane fly (*Nephrotoma suturalis*) Loew. I. The spindle. *Chromosoma* 39:145–73
182. Sanger, J. W. 1975. Presence of actin during chromosomal movement. *Proc. Natl. Acad. Sci. USA* 72: 2451–55
183. Cande, W. Z., Lazarides, E., McIntosh, J. R. 1977. A comparison of the distribution of actin and tubulin in the mammalian mitotic spindle as seen by indirect immunofluorescence. *J. Cell Biol.* 72:552–67
184. Schroeder, T. E. 1970. The contractile ring I. Fine structure of dividing mammalian (HeLa) cells and the effects of cytochalasin B. *Z. Zellforsch.Mikrosk.Anat.*109:431–49
185. Schroeder, T. E. 1973. Actin in dividing cells: contractile ring filaments bind heavy meromyosin. *Proc. Natl. Acad. Sci. USA* 70: 1688–92
186. Sanger, J. W. 1975. Changing patterns of actin localization during cell division. *Proc. Natl. Acad. Sci. USA* 72:1913–16
187. Albrecht-Buehler, G. 1977. Daughter 3T3 cells. Are they mirror images of each other. *J. Cell Biol.* 72:595–603

187a. Mabuchi, I., Okuno, M. 1977. The effect of myosin antibody on the division of starfish blastomeres. *J. Cell Biol.* 74:251–63

188. Adelstein, R. S., Pollard, T. D., Kuehl, W. M. 1971. Isolation and characterization of myosin and two myosin fragments from human blood platelets. *Proc. Natl. Acad. Sci. USA* 68:2703–7
189. Booyse, F. M., Hoveke, T. P., Rafelson, M. E. Jr. 1973. Human platelet actin. Isolation and properties. *J. Biol. Chem.* 248:4083–91
190. Booyse, F. M., Hoveke, T. P., Zschocke, D., Rafelson, M. E. Jr. 1971. Human platelet myosin. Isolation and properties. *J. Biol. Chem.* 246:4921–27
191. Webber, A. J., Budtz-Olsen, O. E. 1974. Thrombin-stimulated release

of platelet microfibrils. *Br. J. Haematol.* 26:255–57

192. Kay, M. M. B., Fudenberg, H. H. 1973. Inhibition and reversal of platelet activation by cytochalasin B or colcemid. *Nature* 244:288–89

193. Pollard, T. D., Fujiwara, K., Handin, R., Weiss, G. 1977. Contractile proteins in platelet activation and contraction *Ann. NY Acad. Sci.* 283:218–36

193a. Nachmias, V., Sullender, J., Asch, A. 1977. Shape and cytoplasmic filaments in control and lidocaine-treated platelets. *Blood* 50:39–53

194. Cohen, I., DeVries, A. 1973. Platelet contractile regulation in an isometric system. *Nature* 246:36–37

195. Mukherjee, T. M., Staehelin, L. A. 1971. The fine structural organization of the brush border of intestinal epithelial cells. *J. Cell Sci.* 8:573–99

196. Tilney, L. G., Mooseker, M. 1971. Actin in the brush border of epithelial cells of the chicken intestine. *Proc. Natl. Acad. Sci. USA* 68: 2611–15

197. Booth, A. G., Kenny, A. J. 1976. Proteins of the kidney microvillus membrane. Identification of subunits after sodium dodecyl sulphate/polyacrylamide-gel electrophoresis. *Biochem. J.* 159:395–407

198. Mooseker, M. S. 1976. Brush border motility. Microvillar contraction in triton-treated brush borders isolated from intestinal epithelium. *J. Cell Biol.* 71:417–33

199. Rodewald, R., Newman, S. B., Karnovsky, M. J. 1976. Contraction of isolated brush borders from the intestinal epithelium. *J. Cell Biol.* 70:541–54

199a. Mak, K. M., Trier, J. S., Serfillip, D., Donaldson, R. M. Jr. 1974. Resistance of adult mammalian intestinal mucosa to cytochalasin B. *Exp. Cell Res.* 86:325–32

200. Burgess, D. R., Grey, R. D. 1974. Alterations in morphology of developing microvilli elicited by cytochalasin B. Studies of embryonic chick intestine in organ culture. *J. Cell Biol.* 62:566–74

201. Lewis, M. R. 1923. Reversible gelation in living cells. *Johns Hopkins Hosp. Bull.* 24:373–79

202. Reaven, E. P., Axline, S. G. 1973. Subplasmalemmal microfilaments and microtubules in resting and phagocytizing cultivated macrophages. *J. Cell Biol.* 59:12–27

203. Hartwig, J. H., Davies, W. A., Stossel, T. P. 1977. Contractile protein translocation in macrophage spreading and secretion. *J. Cell Biol.* 75:956–67

204. Allison, A. C., Davies, P., De Petris, S. 1971. Role of contractile microfilaments in macrophage movement and endocytosis. *Nature New Biol.* 232:153–55

205. Wagner, R., Rosenberg, M., Estensen, R. 1971. Endocytosis in Chang liver cells. Quantitation by sucrose-^{3}H uptake and inhibition by cytochalasin B. *J. Cell Biol.* 50:804–17

206. Franke, W. W., Lüder, M. R., Dartenbeck, J., Zerben, H., Keenan, T. W. 1976. Involvement of vesicle coat material in casein secretion and surface regeneration. *J. Cell Biol.* 69:173–95

207. Orci, L., Gabbay, K. H., Malaisse, W. J. 1972. Pancreatic beta-cell web: its possible role in insulin secretion. *Science* 175:1128–30

208. Butcher, F. R., Goldman, R. H. 1974. Effect of cytochalasin B and colchicine on alpha-amylase release from rat parotid tissue slices. Dependence of the effect on N^6, O^2-dibutyryl adenosine 3′,5′-cyclic monophosphate concentration. *J. Cell Biol.* 60:519–23

209. Zurier, R. B., Hoffstein, S., Weissman, G. 1973. Cytochalasin B: effect on lysosomal enzyme release from human leukocytes. *Proc. Natl. Acad. Sci. USA* 70:844–48

210. Ostlund, R. E., Pastan, I. 1976. The purification and quantitation of myosin from cultured cells. *Biochim. Biophys. Acta* 453:37–47

211. Johnson, D. H., McCubbin, W. D., Kay, C. M. 1977. Isolation and characterization of a myosin-like protein from bovine adrenal medulla. *FEBS Lett.* 77:69–74

212. Abercrombie, M. M., Ambrose, E. J. 1962. The surface properties of cancer cells: a review. *Cancer Res.* 22:525–48

213. Ponten, J. 1976. The relationship between in vitro transformation and tumor formation in vivo. *Biochim. Biophys. Acta* 458:397–422

214. Pollack, R., Rifkin, D. 1975. Actin-containing cables within anchorage-dependent rat embryo cells are dissociated by plasmin and trypsin. *Cell* 6:495–505

215. Pollack, R., Osborn, M., Weber,

K. 1975. Patterns of organization of actin and myosin in normal and transformed cultured cells. *Proc. Natl. Acad. Sci. USA* 72:994–98

216. Wang, E., Goldberg, A. R. 1976. Changes in microfilament organization and surface topography upon transformation of chick embryo fibroblasts with Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* 73: 4065–69

217. Altenberg, B. C., Somers, K., Steiner, S. 1976. Altered microfilament structure in cells transformed with a temperature-sensitive transformation mutant of murine sarcoma virus. *Cancer Res.* 36:251–57

218. McNutt, N. S., Culp, L. A., Black, P. H. 1973. Contact inhibited revertant cell lines isolated from SV40-transformed cells IV. Microfilament distribution and cell shape in untransformed, transformed, and revertant Balb/c 3T3 cells. *J. Cell Biol.* 56:412–28

219. Edelman, G. M., Yahara, I. 1976. Temperature-sensitive changes in surface modulating assemblies of fibroblasts transformed by mutants of Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* 73:2047–51

220. Ash, J. F., Vogt, P. K., Singer, S. J. 1976. Reversion from transformed to normal phenotype by inhibition of protein synthesis in rat kidney cells infected with a temperature-sensitive mutant of Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* 73:3603–7

221. Malech, M. L., Lentz, T. L. 1974. Microfilaments in epidermal cancer cells. *J. Cell Biol.* 60:473–82

222. McNutt, N. S. 1976. Ultrastructural comparison of the interface between epithelium and stroma in basal cell carcinoma and control human skin. *Lab. Invest.* 35:132–42

223. Toh, B. H., Muller, H. K. 1975. Smooth muscle-associated antigen in experimental cutaneous squamous cell carcinoma, keratoacanthoma, and papilloma. *Cancer Res.* 35:3741–45

224. Gabbiani, G., Csank-Brassert, J., Schneeberger, J.-C., Kapanci, Y., Trenchev, P., Holborow, E. J. 1976. Contractile proteins in human cancer cells. Immunofluorescent and electron microscopic study. *Am. J. Pathol.* 83:457–65

225. Fine, R. E., Taylor, L. 1976. Decreased actin and tubulin synthesis in 3T3 cells after transformation by SV40 virus. *Exp. Cell Res.* 102: 162–68

226. Wickus, G., Gruenstein, E., Robbins, P. W., Rich, A. 1975. Decrease in membrane-associated actin of fibroblasts after transformation by Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* 72:746–49

227. Armstrong, P. B., Lackie, J. M. 1975. Studies on intercellular invasion in vitro using rabbit peritoneal neutrophil granulocytes (PMNS) *J. Cell Biol.* 65:439–62

228. Willingham, M. C., Yamada, K. M., Yamada, S. S., Pouyssegur, J., Pastan, I. 1977. Microfilament bundles and cell shape are related to adhesiveness to substratum and are dissociable from growth control in cultured fibroblasts. *Cell* 10:375–80

229. O'Neill, F. J. 1975. Selective destruction of cultured tumor cells with uncontrolled nuclear division by cytochalasin B and cytosine arabinoside. *Cancer Res.* 3111–15

230. Pepe, F. A. 1971. Structural components of the striated muscle fibril. See Ref. 12, pp. 323–53

Ann. Rev. Med. 1978. 29:459–73
Copyright © 1978 by Annual Reviews Inc. All rights reserved

GENETIC TRANSMISSION OF SCHIZOPHRENIA

◆7291

Dennis K. Kinney, Ph.D.
Department of Psychology, University of California, Los Angeles, California 90024

Steven Matthysse, Ph.D.[1]
Department of Psychiatry, Harvard University, and Mailman Research Laboratories, McLean Hospital, Belmont, Massachusetts 02178

INTRODUCTION

Several factors make schizophrenia a particularly serious disorder. The lifetime risk of developing schizophrenia is usually estimated to be about 1%, but may be several times greater if one includes borderline cases (1–4). A high suicide rate (5) is but one sign of the suffering associated with schizophrenia. Moreover, onset tends to occur in adolescence or early adulthood, and many schizophrenics are institutionalized most of their adult lives. As Dean (6) has noted, the "special tragedy of schizophrenia is its predilection for youth. It is the greatest single threat in the entire field of medicine to potentially productive young people."

The term "schizophrenia" was coined by Bleuler (7) for a group of disorders whose primary, or essential, features were thought disorder (characterized by illogical ideas and a "loosening" of associations), markedly inappropriate or unresponsive emotional states, extreme ambivalence, and autistic withdrawal from social contacts and reality into an inner fantasy world. Outstanding secondary features included delusions, hallucinations, and bizarre motor behavior. The third edition of the Diagnostic and Statistical Manual of the American Psychiatric Association (8) requires the presence of at least one of four essential features: delusions, hallucinations, formal thought disorder, or catatonia. The symptom picture at one point in time is, however, not sufficient to distinguish schizophrenia from schizophreniform psychoses with quite different onset and course, response to drug treatment and family history (9).

The task of genetic analysis is complicated by individual and national differences in the diagnosis of schizophrenia (concurrence between judges is 50–90%

---

[1] Supported by Research Scientist Development Award, PHS Grants K2-MH-00108, MH-16674, and MH-25515

0066-4219/78/0401-0459$01.00

in most studies) (8, 10) and the absence of standardized, quantitative measures of severity. Still, much significant work has been done on the genetics of schizophrenia. Because this work has been the subject of several comprehensive reviews (3, 11–14), we emphasize recent studies that provide evidence for genetic transmission, suggest promising directions for future research, or have important clinical implications.

CONSANGUINITY AND TWIN STUDIES

The incidence of schizophrenia increases progressively with a relative's degree of genetic relatedness (consanguinity) to a schizophrenic proband (1, 15). The estimated incidence of schizophrenia in the general population is about 1%, as opposed to 12% for offspring with one schizophrenic parent, and 39% for offspring of two schizophrenics. The incidence is approximately 9% in schizophrenics' siblings, but 47% in their identical twins (15). Unfortunately, such studies are inconclusive, because relatives' consanguinity tends to parallel their exposure to similar environmental factors.

Genetically, monozygotic (MZ) twins are identical, whereas dizygotic (DZ) twins are no more alike than ordinary sibs, so all genetic hypotheses require that a higher percentage of MZ than DZ pairs be concordant (i.e. both co-twins show the same trait). Although precise concordance rates vary, depending on proband selection, diagnostic criteria, and subjects' ages, over a dozen studies in the US, Japan, and Europe have consistently found higher concordance rates for MZ than for DZ twins. Recent studies using accurate zygosity determination, representative samples, and reliable diagnoses unbiased by knowledge of co-twin status suggest that the higher MZ concordance is not a methodological artifact (3, 12, 14, 16, 17).

What remains in question is the interpretation of higher MZ concordance. Twins are biologically deviant; they suffer high prematurity and infant mortality rates (18–21). However, neither the incidence nor the types of schizophrenia in twins differ from those in nontwins (22). MZ twins are more likely than DZ twins to identify with and learn from one another, be dressed and treated alike—indeed, to share most childhood experiences (23); this fact might explain their higher concordance.

Twins reared apart provide crucial evidence on this point, for concordance of MZ twins reared apart is quite similar to those reared together. This holds true even when separated and nonseparated twins are ascertained from the same populations and diagnosed by the same investigators (17). There are difficulties in interpreting these data, however: The number of pairs is small, twins were reared apart for differing periods, the amount of contact between twins while reared by different caretakers is not always clear, and selection may operate in deciding which twin to place for adoption. Most important, more similar environments before or during birth could account for the higher concordance in MZ than DZ twins (24).

Obstetrical complications (OCs) are significantly more common in schizophrenics than in matched nonschizophrenic controls, including their own well sibs or identical co-twins. A recent scholarly review concludes that OCs do not appear to be more common in patients with schizophrenic mothers than in patients with schizophrenic fathers or with normal parents (25). These findings suggest OCs contribute to, rather than result from, schizophrenia.

MZ and DZ twins differ markedly in the frequency of some OCs. Since no DZ but most (70%) MZ twins share one chorion, MZ twins have a greater risk of fetal hypoxia. MZ twins suffer more than DZ twins from prematurity, low birth weight, and perinatal death, and MZ twins are more concordant for these particular OCs (24, 26). However, if these OCs are responsible for the greater schizophrenic concordance of MZ twins, it is hard to explain why the incidence of schizophrenia is no greater for MZ than DZ twins (22, 26).

Boklage (27) has reported a striking association in MZ twins between concordance for schizophrenia and concordance for handedness: Of 12 MZ pairs in which both twins were right-handed, 11 (92%) were concordant for schizophrenia. By contrast, of the 16 pairs in which at least one twin was left-handed, only 4 (25%) were concordant for schizophrenia. Boklage suggests that the schizophrenia of concordant MZ twins may typically be a manifestation of a highly penetrant gene(s), whereas the disorder in discordant MZ twin pairs may be caused by brain injury or disturbance in normal brain lateralization. This association has yet to be confirmed in another sample.

In summary, twin studies provide further support, but not conclusive proof, for genetic transmission of schizophrenia. Ironically, the only definitive conclusion the twin data allow is that environmental variables are important causal factors in schizophrenia—a conclusion demanded by the fact that far less than 100% of MZ twin pairs are concordant for schizophrenia.

ADOPTION STUDIES

The best evidence for genetic causation of schizophrenia comes from adoption studies, which permit one to distinguish the influence of genetic factors from the postnatal environment. Because of its concern for distinguishing genetic influences from those of the pre- and peri- as well as the postnatal environment, particular importance is attached to the research of Kety, Rosenthal, Wender, Schulsinger & Jacobsen in Denmark (5, 28–31). By using the unusually complete social register information available in Denmark, Kety et al were able to identify 33 pairs of schizophrenic and control adoptees matched for age, sex, socioeconomic status of adoptive parents, and age at placement with adoptive parents. These adoptees constituted the probands for the study. Over 90% of the living biological and adoptive parents and sibs of these probands consented to an extensive psychiatric interview, which permitted a blind consensus diagnosis by three raters.

The results strongly support the operation of genetic factors in the transmission of schizophrenia, for they show significantly more schizophrenia among the biological relatives of the schizophrenic probands than among either their adoptive relatives or among the biological or adoptive relatives of the control probands. Chronic schizophrenia was diagnosed among 2.9% of the schizophrenic probands' biological relatives vs 0.6% of the other three groups of relatives. Definite or possible schizophrenia appeared in 13.9% of the schizophrenic probands' biological relatives, but in only 2.7% of their adoptive relatives and 3.4% of the biological relatives of the control probands.

Explanations of the results as artifacts seem unlikely, for precautions were taken to insure unbiased diagnosis, and most adoptees had little or no contact with their biological relatives. Moreover, similar results were found when diagnoses of relatives were based upon institutional records rather than interviews, or when only adoptees separated from their biological parents within one month after birth were included (28–30).

Most persuasive are data on 127 biological paternal half-siblings of the schizophrenic and control adoptees: Significantly more schizophrenia was found among the biological half-sibs of the schizophrenic adoptees. These paternal half-siblings shared neither the same uterus, neonatal mothering experience, nor postnatal environment with the adoptees. Moreover, if pre- or perinatal factors explained the excess schizophrenia among the biological half-sibs of adopted schizophrenics, then one might expect proportionately more schizophrenia among maternal than paternal half-sibs. This is not the case (5, 29).

Kety et al convincingly rebutted (31, 32) critiques (33, 34) of this study. Moreover, the conclusions are supported by other adoption studies using complementary strategies (4, 35–38). In both American (36) and Danish (37) samples, schizophrenia was more prevalent in adoptees with a schizophrenic biological parent than in matched control adoptees whose biological parents were not known to be psychotic. This was true even though most adoptees were separated from their biological parents in early infancy, and (in the Danish sample) the schizophrenic parents included fathers as well as mothers. In a cross-fostering study, adoptees with a schizophrenia-spectrum biological parent but normal adoptive parents had more psychopathology than matched adoptees with normal biological and normal adoptive parents. By contrast, cross-fostered adoptees with normal biological parents but schizophrenia-spectrum adoptive parents did not show more psychopathology than the controls (38).

In summary, adoption studies indicate that the tendency for schizophrenia to run in families reflects genetic factors. They suggest, furthermore, that schizophrenia in the rearing family is not an important etiological factor. An influence of the psychosocial environment is not excluded, however, since schizophrenogenic parents need not necessarily be schizophrenic.

MODELS OF TRANSMISSION

Schizophrenia is not a simple Mendelian character, although if the disorder is heterogeneous (see p. 465), it is possible that one or more of the subtypes is Mendelian. The characteristic segregation ratios of Mendelian inheritance would become blurred if mixed in with other types. Most genetic theorizing, however, has assumed that schizophrenia is all of one type, and has attempted to find the best-fitting model of the common mode of inheritance. The two most commonly discussed hypotheses are the single major locus and polygenic models. The single major locus model is more flexible than classical dominant or recessive transmission, because it includes incomplete penetrance (not all gene carriers, even if homozygous, need to manifest the trait), phenocopies (non-gene carriers who nevertheless have the trait), and intermediate dominance (the probability of expression in heterozygotes can be in between that of genetically normal and homozygous individuals, or even outside the interval between them). The polygenic model postulates a very large number of genes, each one of similar but small effect; the effects are assumed to be additive, so that the distribution of overall genetic liability follows a normal curve. An indefinite number of variants of these models are possible. There may be two genes, interacting in such a way that both must be present to produce schizophrenia (epistasis). Or there may be a very large number of loci that contribute to the risk but dissimilar in magnitude, some behaving like major genes and others like minor genes. If the parameters of the models are chosen to fit the observed frequencies of schizophrenia both in probands' offspring and sibs of affected and in the general population, the single major locus and polygenic models make predictions that are difficult to distinguish in practice. The frequency in monozygotic twins tends to be overestimated by the polygenic model and underestimated by the single major locus model (15). The fact that these discrepancies lie in opposite directions suggests that the most appropriate model might lie in between: several loci, or a polygenic model of the mixed major gene–minor gene type. However, the variety of possible constructions fitting these specifications has dissuaded theorists from working them out in detail.

Models of the transmission of schizophrenia serve a useful purpose even though the available data do not permit discriminating between them. Casting the genetic ideas into mathematical form helps to formulate them precisely and also suggests certain research strategies that might otherwise be overlooked. In the case of the single major locus model, the proportions of genetically normal, heterozygous, and homozygous individuals in the population depend on the gene frequency q according to the Hardy-Weinberg law: $(1-q)^2$, $2q(1-q)$, q^2. The proportion of each genetic type that is affected depends on the penetrances f_0, f_1, f_2, which are determined by q in order to fit the data on incidence in offspring and sibs of affected and in the general population. From the genotype frequencies and the penetrances, one may deduce for each possible gene frequency the proportion of schizophrenics that are genetically normal (phenocopies), heterozygous, and homozygous. The actual gene frequency is not uniquely determined by the incidence data, but can be within a certain range. If the gene frequency is very low, the majority of schizophrenics will be phenocopies and only one in 769 will be a homozygote. If the gene frequency is high, there will be no phenocopies, and one in 45 will be a homozygote. The bulk of the schizophrenic population will be heterozygotes (15).

Certain conclusions may be drawn. First, the possibility exists that many schizophrenics, even the majority, have no genetic basis for their disorder, even though the adoption studies demonstrate that, on the average, there is an inherited tendency. Second, homozygous schizophrenics are rare for any acceptable gene frequency. A random sample from a clinic or hospital population is likely to miss them entirely. Calculations based on the polygenic model lead to rather similar conclusions: only 9.1% of the schizophrenic population have a gene concentration so high that, were they to start life over in a new environment, they would have a 99% chance of again becoming schizophrenic.

Any modern discussion of models of transmission of schizophrenia must take into account the possibility of cultural inheritance, outlined in a classic paper by Cavalli-Sforza & Feldman. They stated, "When the parents' phenotypes affect the phenotype of their children directly, a new mode of inheritance arises which operates side by side with strictly biological inheritance (through the DNA) and may be difficult to distinguish from it" (39). Cultural transmission in schizophrenia might occur if the rearing practices of schizophrenics tended to produce schizophrenia in their children (the cross-fostering data suggest that this mechanism does not occur—see above); if there were some psychological trait (e.g. disordered communication) that predisposed one to schizophrenia and was handed down from generation to generation; or if some sociological or biological aspect of the environment, similarly transmitted, was a factor in vulnerability (e.g. poverty, diet).

When genetic analyses are carried out on suitable large pedigrees, comprising many individuals in several generations, the possibilities for testing models are somewhat better. First, it is likely that only one mechanism of transmission is present in any one family. Second, analytical techniques have been developed for comparing the likelihood that a given pattern of affected and unaffected relatives will occur, if the underlying mechanism is in accord with major locus or polygenic hypotheses (40–42). These methods have become very sophisticated mathematically (43, 44), but it is well to keep certain limitations in mind: (*a*) The method of sampling the pedigree affects the likelihoods, and corrections are necessary to prevent artifacts (45). (*b*) Even though a model predicts such a small likelihood that it can be excluded statistically, the mode of transmission that model represents cannot be immediately discarded. In any calculation, one is testing not only a particular genetic mechanism, but also certain assumptions which may not be correct, like random mating and Hardy-Weinberg equilibrium. (*c*) Individuals with doubtful diagnoses, or who are unaffected but still within the age of risk, pose problems, although some of the recent analytical methods can accommodate these factors. (*d*) Consistency with a model is only meaningful if the analytical method is powerful enough, with the sample size available, to rule out alternative models. Studies of the statistical power of segregation analysis have been carried out by computer simulation (46); the power to detect a major locus in the presence of common environmental factors shared by siblings, using only two-generation families, was not impressive. These calculations need to be extended to multigenerational pedigrees as well. The power of pedigree analysis without adoption to distinguish between genetic and cultural transmission has also not been tested. Preliminary results using only frequencies in relatives show that some cultural transmission models do indeed mimic genetic transmission (47).

It is difficult to reconcile a relatively high and apparently stable incidence of schizophrenia with schizophrenics' low fertility, which should exert strong selection pressure against genes increasing vulnerability to schizophrenia. Such genes could be maintained if, as in sickle-cell anemia in regions where malaria is endemic, unaffected carriers gain a compensatory advantage in increased fertility or survival through the reproductive period (48). A rigorous study (49) using matched controls and blind diagnoses confirmed prior reports of lower fertility in schizophrenics, but found no reproductive advantage for their biological sibs. Some reports suggest a basis for increased survival in carriers [e.g. fewer viral infections in schizophrenics and their relatives (50)], but these await replication. Kidd (51) showed that, for a single gene model, schizophrenia could be maintained at its present rate by an increased fitness of 5% or less in unaffected carriers; for practical purposes, such a small heterozygote advantage might be undetectable.

In conclusion, understanding the mode of transmission of schizophrenia will probably have to await delineation of biological and psychological risk factors and resolution of the illness into homogeneous subtypes. Until then, concepts like the schizophrenic gene or genetic liability are best not taken too literally.

HETEROGENEITY

Several lines of evidence indicate that schizophrenia is etiologically heterogeneous.[2] Clinical subtypes not only differ in symptoms, signs, course, and outcome, but also in risk of schizophrenia to probands' family members. Subtypes (e.g. poor-prognosis schizophrenia, especially hebephrenia) tend to be consistent within families and to be concordant in MZ twins (17, 57, 58).

The risk estimates for schizophrenia in relatives of a schizophrenic proband (see previous discussion) (11) were derived from probands with classical chronic schizophrenia having insidious onset and deteriorating course. For borderline and latent schizophrenic probands, the risk in relatives is likely to be considerably lower, and the risk is also lower in relatives of paranoid as compared with hebephrenic patients (57, 58). Indeed, in the case of acute schizophreniform psychoses with a good premorbid history and a remitting course, there may be an increased risk of manic-depressive disorder rather than schizophrenia in the relatives. This may be true even if florid schizophrenic symptoms are present in the proband (9). Conversely, manic-depressives can manifest phases that are atypical yet definitely schizophreniform (59). In the Danish adoption study (5) there were extremely few cases of acute schizophrenic reaction in biological relatives of chronic schizophrenic adoptees; conversely, little schizophrenia was seen in biological relatives of adoptees diagnosed as acute schizophrenics.

The adoption study data support the concept of a schizophrenia spectrum of genetically related disorders, ranging from the most severe and chronic cases of schizophrenia, through milder cases with borderline, latent, or uncertain symptoms (5, 29, 30). These disorders, like chronic schizophrenia, were more often found among the biological relatives of schizophrenic than control probands.

---

[2] A similar situation prevails in affective disorders, where there appear to be low serotonin and low noradrenaline subgroups with different clinical characteristics (52) and drug response (53–55). In some pedigrees the illness seems to be X-linked; in others it does not (56).

Neither definite nor uncertain schizophrenia was found in the biological relatives of 16 of the 33 schizophrenic adoptees. It is well established that schizophrenia-like symptoms can result from nongenetic causes, as in the psychogenic psychoses and organic brain syndromes (60, 14). A recent review of heterogeneity among schizophrenic patients concludes that "little remains to be gained from studies in which heterogeneous samples of schizophrenic patients are contrasted with matched samples of nonschizophrenic control subjects" (61).

Intensive study of a single large family is more likely to isolate a subtype of schizophrenia in which a single etiology is involved. A pedigree with several affected individuals should be especially fruitful, because the information yielded by a pedigree increases sharply with the number of informative family members. Such high density pedigrees would have the additional advantage of garnering a greater concentration of cases homozygous for an allele predisposing to schizophrenia, although a high concentration of environmental predisposing factors may also be present. Homozygotes might manifest biochemical or other physiological traits that are more markedly deviant than in heterozygotes. Frequently, biological studies on randomly sampled schizophrenics reveal no statistically significant difference between the groups, but a few individuals appear to be markedly deviant from the norm. If such outliers had pedigrees with multiple cases, study of their family members would be particularly worthwhile. Rieder & Gershon (62) proposed some ingenious strategies for using pedigree methods to clarify etiological factors.

BIOLOGICAL RISK FACTORS

The two most widely discussed biological correlates of schizophrenia—platelet monoamine oxidase (MAO) and smooth-pursuit visual tracking—are believed to be heritable. The correlation in platelet MAO between monozygotic twins was 0.88, compared with 0.45 between dizygotic twins (63). Low platelet MAO is characteristic of chronic schizophrenia, not acute schizophreniform psychoses (63), but it is also present in bipolar-affective-disorder patients and their first-degree relatives (64). Since nonschizophrenic MZ twins of schizophrenics, and well relatives of bipolar-affective-disorder patients, also tend to have low platelet MAO, the trait is thought to be an indicator of genetic vulnerability to these disorders (63, 64). The results with smooth-pursuit visual tracking are similar. The correlation in velocity-arrest scores was 0.77 between MZ twins and 0.40 between DZ twins. Since 4 of 9 of the nonschizophrenic MZ twins of schizophrenics showed deviant tracking, this trait also may be an indicator of genetic vulnerability (65).

Buchsbaum et al (66) devised a very useful approach to studying the specificity of these biological correlates by inverting the usual research design.

Instead of starting with schizophrenic and normal groups and comparing MAO, they selected the upper and lower deciles of MAO and compared psychopathology. The low-MAO subjects ascertained in this way were not schizophrenic. On the average, they had more psychiatric contact and hospitalization, more criminal behavior, and more suicide attempts. Suicide attempts were also more prevalent in their families. Consequently it is more reasonable to regard low platelet MAO as a correlate of generalized psychiatric vulnerability than of schizophrenia. The same approach has not yet been applied to visual tracking.

The broad spectrum of the psychopathology associated with low platelet MAO suggests that searching for heritable risk factors may be a more useful genetic strategy than hoping to find a genome that codes directly for schizophrenia. Childs (67) convincingly defended this approach by analogy with other medical conditions, such as early myocardial infarction for which several risk factors are known: hyperlipidemia, hypertension, diabetes, obesity, and smoking. The individual risk factors may have both genetic and environmental components. Childs concludes: "The genetic contribution to heart attacks will differ in different families so that each man's coronary occlusion may have slightly different genetic causes. Thus, it is not coronary occlusion that is inherited, but only one or a number of genes whose actions fail to support homeostasis in the face of special experiences. There should be some lessons here for the detailed analysis of genetic contributions to psychiatric disease" (67).

If there are several genetically distinct risk factors for schizophrenia, there may be different environmental potentiators for each. The observation of Mishler & Waxler (68) that good premorbid schizophrenics have abnormal family interaction patterns, whereas poor premorbid patients do not, is consistent with this idea.

Association between certain HL-A antigens and schizophrenia has been found. According to the most thorough of these studies, the antigen A-28 was found in 18.9% of schizophrenics compared with 6.3% of controls ($P < .01$, after correction for the number of antigens tested) (69). Other associations have also been reported: an elevated frequency of A-9 (70) and a decrease of A-10 (71). Curiously, although all these results were statistically significant after correction for the number of antigens tested, they each disagreed with the others. Differences in methodology, or genetic diversity in the populations tested, may account for these differences. Although an HL-A association would be of obvious genetic interest (most associations in the HL-A region are the result of closely linked genes that have remained in a state of linkage disequilibrium), the results need confirmation with independent samples.

Conversely, the discovery of linkage to a genetic marker (blood group, enzyme polymorphism) would be strong evidence for a major gene. Although in some cases, such as the histocompatibility loci, association and linkage may be related, they are different concepts. Linkage refers to physical proximity on the chromosome, which reveals itself by nonindependent assortment of the genes within pedigrees; either coupling (positive association) or repulsion (negative association) of the two linked traits within the pedigree may be observed. The large number of mapped genes on the human chromosomes makes linkage detection relatively feasible, when a major locus is known to be operating (72–74). A marker is likely to be useful for linkage analysis only if its distance is less than about 20 recombination units from the trait locus; otherwise the number of siblings required is too large. The probability of at least one of 30 markers lying within 20 recombination units of an arbitrary trait locus is approximately 30%. The feasibility of linkage analysis is less evident in schizophrenia, where a major locus is not known to be involved. Clarification of the problem of heterogeneity, and classification of independently transmitted risk factors, would make the task of detecting linkage much easier.

ENVIRONMENTAL RESEARCH STRATEGIES

In order to show that an environmental correlate such as abnormal behavior patterns in the family (75) is a cause of schizophrenia (hypothesis *D*), one must first exclude three alternative hypotheses, *A*, *B*, and *C*. Hypothesis *A*: The schizophrenic's symptoms and the abnormal behavior of his parent(s) both result from pathogenic genes, which the schizophrenic and his parent(s) share through inheritance. It has been suggested that early parental death is a causative factor in schizophrenia, but among adopted schizophrenics an excess of early deaths was observed only in the biological and not in the adoptive parents, which raises the possibility that the early deaths of the parents and the schizophrenic illness of the offspring both result from shared genes (5). Hypothesis *B*: The abnormal parental behavior may be an artifact of the test or interview situation, which causes the parents of schizophrenics to manifest unrepresentative behavior (e.g. from anxiety). Schopler & Loftin (76) found that greater indications of thought disorder in parents of childhood schizophrenics possibly resulted from anxiety created by discussion of their schizophrenic child's problems. Hypothesis *C*: The parents' abnormal behavior is the effect of rearing a schizophrenic offspring. Liem (77) found that impairment on a laboratory measure of accuracy in communication was characteristic of control parents as well as parents of schizophrenic children, when their task was to communicate with schizophrenics. Early parental death, thought disorder, and deviant communication may indeed be contributing factors, but the methods used to study them have not been conclusive.

Wynne, Singer & Toohey (78) recently strengthened their hypothesis that parental communication deviance is a cause of schizophrenia: They demonstrated that it was more common in the adoptive parents of schizophrenics than controls. This strategy can rule out Hypothesis *A*, but it cannot exclude Hypotheses *B* or *C*. Similarly, environmental factors that distinguish MZ twins discordant for schizophrenia (e.g 79) can exclude Hypothesis *A* (since genetic factors are held constant within each twin pair), but not Hypotheses *B* and *C*. Even prospective studies of children at high risk cannot definitively exclude Hypotheses *B* and *C*, because environmental correlates might be effects of deviant behavior present in the child before he manifests diagnosable schizophrenia.

There is a research design that in principle is capable of deciding between Hypothesis *D* and the other three alternatives: comparison of the environments of schizophrenic adoptees at high genetic risk for schizophrenia with those at low genetic risk (80). Environmental causes should be especially pronounced in cases where there is little or no hereditary predisposition. The incidence of schizophrenia among the adoptees' biological relatives can be used to estimate their genetic risk. Because all adoptees being compared are schizophrenic, environmental factors inversely correlated with genetic risk are likely to be causes rather than effects of schizophrenia.

GENETIC COUNSELING

The evidence from adoption, twin, and consanguinity studies provides a justification for genetic counseling of schizophrenic patients and their families (81). The risk to relatives can be estimated in simple cases on the basis of empirical incidence figures (see p. 460), even though the mode of transmission is unknown. In cases where the family history presents a complex pattern of psychiatric illnesses in several generations, exact estimates of risk cannot be made without knowledge of the genetic mechanism. In the practical counseling situation, precise risk figures are often less crucial than assessment of the subjective burden that the risk entails.

One of the major problems in applying risk figures to individual cases is ambiguity of diagnoses. As mentioned above, the risk for schizophrenia to relatives of borderline schizophrenics and patients with acute schizophreniform psychoses is likely to be lower than in chronic schizophrenia, although in the case of acute psychosis there may also be a risk of manic-depressive illness in relatives. At present there is little concrete advice that can be given the prospective parent regarding minimizing the risk. If indeed there are independent biological risk factors, each with its own environmental precipitants, their clarification will make possible a much more rational approach to counseling and prevention.

SUMMARY

There is convincing evidence from consanguinity, twin, and especially adoption studies that schizophrenia has a genetic component. There is also evidence from studies of MZ twins for an environmental component. It is not known whether the relevant environmental factors are prenatal or postnatal, psychosocial or physical; the presence of schizophrenia in the rearing family does not seem to be etiologically significant. It is likely that schizophrenia, especially if defined by symptoms alone, is etiologically heterogeneous. Study of pedigrees with several affected individuals is a useful approach to resolving this heterogeneity. The relevant genes may not code directly for schizophrenia, but for risk factors that predispose to the illness under particular environmental conditions. Careful methodology is needed to distinguish cause from effect in studying psychosocial factors. Genetic counseling is possible on the basis of empirical risk figures, but clarification of biological risk factors and environmental precipitants will make possible a much more rational approach to counseling and prevention.

Literature Cited

1. Kidd, K. K., Cavalli-Sforza, L. L. 1973. An analysis of the genetics of schizophrenia. *Soc. Biol.* 20:254–65
2. Rosenthal, D., Wender, P. H., Kety, S. S., Welner, J., Schulsinger, F. 1971. The adopted-away offspring of schizophrenics. *Am. J. Psychiatry* 128:307–11
3. Rosenthal, D. 1970. *Genetic Theory and Abnormal Behavior.* New York: McGraw-Hill. 317 pp.
4. Wender, P. H. 1972. Adopted children and their families in the evaluation of nature-nurture interactions in the schizophrenic disorders. *Ann. Rev. Med.* 23:355–71
5. Kety, S. S., Rosenthal, D., Wender, P. H., Schulsinger, F., Jacobsen, B. 1978. The biological and adoptive families of adopted individuals who became schizophrenic: prevalence of mental illness and other characteristics. In *Schizophrenia: Nature and Treatment,* ed. L. D. Wynne, R. L. Cromwell, S. Matthysse. New York: Wiley. In press
6. Dean, S. R. 1975. *Schizophrenia: The First Ten Dean Award Lectures,* p. 314 New York: MSS Information Corp.
7. Bleuler, E. 1911.*Dementia Praecox or the Group of Schizophrenias.* Transl. J. Zinkin. 1950. New York: Int. Univ. Press. 548 pp.
8. Spitzer, R. L., Andreasen, N., Endicott, J., Woodruff, R. A. 1978. Proposed classification of schizophrenia in DSM-III. See Ref. 5
9. Pope, H., Lipinski, J. 1977. Differential diagnosis of schizophrenia and manic depressive illness. *Arch. Gen. Psychiatry.* In press
10. Falek, A., Hanna, M. M. 1975. Classification in schizophrenia. *Arch. Gen. Psychiatry* 32:59–67
11. Slater, E. 1968. A review of earlier evidence on genetic factors in schizophrenia. In *The Transmission of Schizophrenia,* ed. D. Rosenthal, S. S. Kety, pp. 15–26. Oxford: Pergamon. 433 pp.
12. Gottesman, I. I., Shields, J. 1972. *Schizophrenia and Genetics, A Twin Study Vantage Point.* New York: Academic. 433 pp.
13. Kaplan, A. R., ed. 1972. *Genetic Factors in Schizophrenia.* Springfield, Ill.: Thomas. 698 pp.
14. Worden, F. G., Childs, B., Matthysse, S., Gershon, E. S., eds. 1976. Frontiers of psychiatric genetics. *Neurosci. Res. Program. Bull.* 14:8–86
15. Matthysse, S. W., Kidd, K. K. 1976. Estimating the genetic contribution to schizophrenia. *Am. J. Psychiatry* 133:185–91
16. Kringlen, E. 1968. An epidemiological-clinical twin study on schizophrenia. See Ref. 11, pp. 49–63
17. Inouye, E. 1972. Monozygotic twins with schizophrenia reared apart in infancy. *Jpn. J. Hum. Genet.* 16: 182–90
18. Bulmer, M. G. 1970. *Biology of Twinning in Man.* New York: Oxford Univ. Press. 205 pp.
19. Potter, E. L. 1963. Twin zygosity and placental form in relation to the outcome of pregnancy. *Am. J. Obstet. Gynecol.* 87:566–77
20. Yerulshalmy, J., Sheerar, S. E. 1940. Studies on twins. I. The relation of order of birth and age of parents to the frequency of like-sexed and unlike-sexed twin deliveries. *Hum. Biol.* 12:95–113
21. Yerulshalmy, J., Sheerar, S. E. 1940. Studies on twins. II. On the early mortality of like-sexed and unlike-sexed twins. *Hum. Biol.* 12: 247–63
22. Rosenthal, D. 1960. Confusion of identity and the frequency of schizophrenia in twins. *Arch. Gen. Psychiatry* 3:297–304
23. Koch, H. L. 1966. *Twins and Twin Relations.* Chicago: Univ. Chicago Press. 302 pp.
24. Campion, E., Tucker, G. 1973. A note on twin studies, schizophrenia, and neurological impairment. *Arch. Gen. Psychiatry* 29:460–64

25. McNeil, T., Kaij, L. 1978. Obstetric factors in the development of schizophrenia. See Ref. 5
26. Pollin, W. 1972. The pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* 27:29–37
27. Boklage, C. E. 1977. Schizophrenia, brain asymmetry development, and twinning: cellular relationship with etiological and possibly prognostic implications. *Biol. Psychiatry*. In press
28. Kety, S. S., Rosenthal, D., Wender, P. H., Schulsinger, F. 1968. The types and prevalence of mental illness in the biological and adoptive families of adopted schizophrenics. See Ref. 11, pp. 345–62
29. Kety, S. S., Rosenthal, D., Wender, P. H., Schulsinger, F., Jacobsen, B. 1975. Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic: A preliminary report based upon psychiatric interviews. In *Genetic Research in Psychiatry*, ed. R. Fieve, D. Rosenthal, H. Brill, pp. 147–65. Baltimore: Johns Hopkins Univ. Press. 301 pp.
30. Kety, S., Rosenthatl, D., Wender, P. H., Schulsinger, F., Jacobsen, B. 1976. Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic. *Behav. Genet.* 6:219–25
31. Kety, S. S., Rosenthal, D., Wender, P. H., Schulsinger, F. 1976. Studies based on a total sample of adopted individuals and their relatives: why they were necessary, what they demonstrated and failed to demonstrate. *Schizophr. Bull.* 2:413–28
32. Kety, S. S. 1975. Dr. Kety replies. *Am. J. Psychiatry*. 132:458–59
33. Gottesman, I. I., Shields, J. A. 1976. A critical review of recent adoption, twin, and family studies of schizophrenia: behavioral genetics perspectives. *Schizophr. Bull.* 2:36 -401
34. Litwak, T. 1975. Accuracy of diagnosis and genetic factors in schizophrenia. *Am. J. Psychiatry* 132:457–58
35. Wender, P. H., Rosenthal, D., Kety, S. S. 1968. A psychiatric assessment of the adoptive parents of schizophrenics. See Ref. 11, pp. 235–50
36. Heston, L. L. 1966. Psychiatric disorders in foster home reared children of schizophrenic mothers. *Br. J: Psychiatry* 112:819–25
37. Rosenthal, D., Wender, P. H., Kety, S. S., Schulsinger, F., Welner, J., Ostergaard, L. 1968. Schizophrenics' offspring reared in adoptive homes. See Ref. 11, pp. 377–91
38. Wender, P. H., Rosenthal, D., Kety, S. S., Schulsinger, F., Welner, J. 1974. Crossfostering: a research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia. *Arch. Gen. Psychiatry* 30:121–28
39. Cavalli-Sforza, L. L. and Feldman, M. W. 1973. Cultural versus biological inheritance: phenotypic transmission from parents to children (a theory of the effect of parental phenotypes of children's phenotypes). *Am. J. Hum. Genet.* 25:618–37
40. Elston, R. C., Stewart, J. 1971. A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21:523–42
41. Ott, J. 1974. Estimation of the recombination fraction in human pedigrees. *Am. J. Hum. Genet.* 26:588–97
42. Lange, K., Westlake, J., Spence, M. 1976. Extensions to pedigree analysis. II. Recurrence risk calculations under the polygenic threshold model. *Hum. Hered.* 26:337–48
43. Lange, K., Elston, R. C. 1975. Extensions to pedigree analysis. I. Likelihood calculations for simple and complex pedigrees. *Hum. Hered.* 25:95–105
44. Cannings, C., Thompson, E. A., Skolnick, M. H. 1976. The recursive derivation of likelihoods on pedigrees. *Adv. Appl. Probabil.* 8: 622–25
45. Elston, R. C. 1973. Ascertainment and age of onset in pedigree analysis. *Hum. Hered.* 23:105–12
46. MacLean, C. J., Morton, N. E., Lew, R. 1975. Analysis of family resemblance: IV. Operational characteristics of segregation analysis. *Am. J. Hum. Genet.* 27:365–84
47. Rao, D. C., Morton, N. E. 1974. Path analysis of family resemblance in the presence of gene-environment interaction. *Am. J. Hum. Genet.* 26:767–72
48. Huxley, J., Mayr, G., Osmond, H.,

Hoffer, A. 1964. Schizophrenia as genetic morphism. *Nature* 204:220–21

49. Rimmer, J., Jacobsen, B. 1976. Differential fertility of adopted schizophrenics and their half-siblings. *Psychiatr. Res.* 54:161–66
50. Carter, M., Watts, C. A. H. 1971. Possible advantages among schizophrenics' relatives. *Br. J. Psychiatry* 118:453–60
51. Kidd, K. K. 1975. On the possible magnitudes of selective forces maintaining schizophrenia in the population. See Ref. 29, pp. 135–45
52. Asberg, M., Träskman, L., and Thorén, P. 1976. 5-HIAA in the cerebrospinal fluid: a biochemical suicide predictor? *Arch. Gen. Psychiatry* 33:1193–97
53. Maas, J. W. 1975. Biogenic amines and depression: biochemical and pharmacological separation of two types of depression. *Arch. Gen. Psychiatry* 32:1357–61
54. Van Praag, N. M. 1977. Significance of biochemical parameters in the diagnosis, treatment, and prevention of depressive disorders. *Biol. Psychiatry* 12:101–31
55. Goodwin, F. K., Post, R. M. 1975. Studies of amine metabolites in affective illness and in schizophrenia: a comparative analysis. In *Biology of the Major Psychoses,* ed. D. X. Freedman. Res. Publ. Assoc. Res. Nerv. Ment. Dis., Vol. 54, pp. 299–332. New York: Raven. 372 pp.
56. Gershon, E. S., Bunney, W. E. Jr. 1976. The question of X-linkage in bipolar manic-depressive illness. *J. Psychiat. Res.* 13:99–117
57. Tsuang, M. T. 1975. Heterogeneity of schizophrenia. *Biol. Psychiatry* 10:465–74
58. Winokur, G. 1975. Paranoid vs. hebephrenic schizophrenia: clinical and familial (genetic) heterogeneity. *Psychopharmacology* 1:567–77
59. Stromgren, E. 1965. Schizophreniform psychoses. *Acta Psychiatr. Scand.* 41:483–89
60. Davison, K., Bagley, C. R. 1969. Schizophrenic-like psychoses associated with organic disorders of the central nervous system: a review of the literature. *Br. J. Psychiatry* Spec. Publ. 4:113–84
61. Houlihan, M. P. 1977. Heterogeneity among schizophrenic patients: selective review of recent findings (1970–75). *Schizophr. Bull.* 3:246–58
62. Rieder, R. O., Gershon, E. S. 1977. Genetic strategies: approaches to the etiology of psychiatric disorders. *Arch. Gen. Psychiatry* In press
63. Wyatt, R. J., Belmaker, R., Murphy, D. 1975. Low platelet monoamine oxidase and vulnerability to schizophrenia. In *Modern Problems of Pharmaco-Psychiatry, Vol. 10, Genetics and Psychopharmacology,* ed. E. J. Mendlewicz, pp. 38–57. Basel: Karger. 132 pp.
64. Leckman, J. F., Gershon, E. S., Nichols, A. S., Murphy, D. L. 1977. Reduced MAO activity in first-degree relatives of individuals with bipolar affective disorders. *Arch. Gen. Psychiatry* 34:01–6
65. Holzman, P. S., Kringlen, E., Levy, D. L., Proctor, L. R., Haberman, S. 1978. Smooth pursuit eye movements in twins discordant for schizophrenia. *J. Psychiatr. Res.* In press
66. Buchsbaum, M. S., Coursey, R. D., Murphy, D. L. 1976. The biochemical high-risk paradigm: behavioral and familial correlates of low platelet monoamine oxidase activity. *Science* 194:339–41
67. See Ref. 14, pp. 16–17
68. Mishler, E., Waxler, N. 1968. *Interaction in Families: An Experimental Study of Family Process and Schizophrenia.* New York: Wiley. 436 pp.
69. Ivanyi, D., Zemek, P., Ivanyi, P. 1976. HL-A antigens in schizophrenia. *Tissue Antigens* 8:217–20
70. Eberhard, G., Franzen, G., Low, B. 1975. Schizophrenia susceptibility and HL-A antigen. *Neuropsychobiology* 1:211–17
71. Smeraldi, E., Bellodi, L., Scorza-Smeraldi, R., Fabio, G., Sacchetti, E. 1976. HLA-SD antigens and schizophrenia: statistical and genetical considerations. *Tissue Antigens* 8 191–96
72. Elston, R. C., Lange, K. 1975. The prior probability of autosomal linkage. *Ann. Hum. Genet.* 38:341–50
73. Lange, K., Spence, M. A., Frank, M. B. 1976. Application of the Lod method to the detection of linkage between a quantitative trait and a qualitative marker: a simulation experiment. *Am. J. Hum. Genet.* 28:167–73

74. Morton, N. E. 1955. Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7:277–318
75. Goldstein, M. J., Rodnick, E. H. 1975. The family's contribution to the etiology of schizophrenia: current status. *Schizophr. Bull.* 14:48–63
76. Schopler, E., Loftin, J. 1969. Thought disorders in parents of psychotic children. *Arch. Gen. Psychiatry* 20:174–81
77. Liem, J. H. 1974. Effects of verbal communications of parents and children: a comparison of normal and schizophrenic families. *J. Consult. Clin. Psychol.* 42:438–50
78. Wynne, L. C., Singer, M. T., Toohey, M. L. 1976. In *Schizophrenia 75: Psychotherapy, Family Studies, Research,* ed. J. Jorstad, E. Ugelstad, pp. 413–51. Oslo: Univ. Oslo Press
79. Mosher, L. R., Pollin, W., Stabenau, J. R. 1971. Identical twins discordant for schizophrenia: Neurological findings. *Arch. Gen. Psychiatry* 24:422–30
80. Kinney, D. K., Jacobsen, B. 1978. Environmental factors in schizophrenia: new adoption study evidence and its implications for genetic and environmenetal research. See Ref. 5. In press
81. Kety, S. S., Matthysse, S., Kidd, K. K. 1978. Genetic counseling for schizophrenic patients and their families. In *Controversy in Psychiatry* eds. J. P. Brady, H. K. H. Brodie. Philadelphia: Saunders. In press

Ann. Rev. Med. 1978. 29:475–83
Copyright © 1978 by Annual Reviews Inc. All rights reserved

LEFT VENTRICULAR STIFFNESS

◆7292

Judith A. Wisneski, M.D. and J. David Bristow, M.D.
Veterans Administration Hospital, and Department of Medicine and Cardiovascular Research Institute, University of California, San Francisco, California 94121

INTRODUCTION

Volume and pressure rise together as a cardiac ventricle fills, and the relationship between these variables has received considerable attention in recent years. Diastolic left ventricular pressure is a primary determinant of left atrial, pulmonary venous, and pulmonary capillary pressures, and when elevated can lead to symptoms. The diastolic pressure-volume relationship is therefore of direct significance in patients.

One sequence leading to increased left ventricular (LV) diastolic pressure begins with impaired systolic emptying and is followed by compensatory overfilling, which results in a high diastolic volume. The elevated volume is associated with high diastolic pressure as the ventricle "moves" predictably along a normal pressure-volume curve. However, it is now evident that the diastolic properties of the ventricle change in response to a variety of influences, and the relationship between pressure and volume can be substantially altered. It follows that the ventricle may be less or more difficult to fill, depending on its physical properties as well as the particular volume at which filling takes place. As a consequence, elevated diastolic pressure can occur because the chamber has reduced distensibility and/or is operating at a high volume. Indeed, it is possible to have congestive signs related primarily to decreased distensibility (1–3), and it is now well established that high LV diastolic pressure can be associated with normal volume.

There are multiple factors that may change ventricular stiffness, including the completeness of myocardial relaxation and the composition and thickness of the ventricular wall (4). Although the details of the mechanisms that alter distensibility are not fully understood, some information is available and is reviewed here.

DEFINITIONS

Some definitions and technical considerations are needed to appreciate the importance and limitations of recently published information. Ventricular di-

0066-4219/78/0401-0475$01.00

astolic *compliance* is similar in a qualitative sense to distensibility and is defined as the passive change in volume divided by the associated change in pressure ($\Delta V/\Delta P$). More commonly, the inverse of this relationship is plotted, and represents ventricular volume *stiffness* (5), as displayed in Figure 1. This pressure-volume (*PV*) relationship is curvilinear, and its slope increases as the ventricle fills (6). The slope defines stiffness; its change during filling means that stiffness increases as the ventricular volume increases. This is a difficult concept to deal with, and attempts to normalize such curvilinear data have been made by estimating stiffness at specific pressures (7), producing a stiffness or elastic modulus that is linear.

It is useful to discriminate between the qualitative, intuitive use of terms such as stiffness and compliance and their more strict mathematical definitions. Authors have used the term increased stiffness (or decreased compliance) to mean simply that ventricular diastolic pressure was abnormally high at a given volume, usually at end-diastole. This could occur because of a shift in the position of the *PV* curve with or without a change in slope of the relationship (Figure 1). Only the latter specifically refers to stiffness or compliance in a rigorous sense. In either case, however, an abnormality of the pressure-volume relationship may be of substantial clinical importance. Both the slope and position of the *PV* curve are important to consider.

TECHNICAL CONSIDERATIONS

An impediment to the study of ventricular stiffness in various diseases has been the technical difficulty in obtaining high quality, simultaneous ventricular pressures and volumes. In most studies LV volume has been determined by angiocardiography, and pressures must also be recorded from the chamber. In one of the early, landmark studies of this problem, complete *PV* curves were obtained from pressures measured just prior to angiographic injection for volume

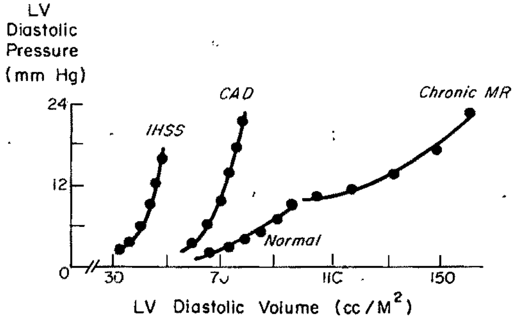


Figure 1 Diastolic pressure-volume relationships in various diseases, adapted from Gaasch and associates (6). Compared with a normal individual there is a shift to the left in the pressure-volume curve in idiopathic hypertrophic subaortic stenosis and coronary artery disease. In sharp contrast, a shift to the right is seen in the chronically dilated ventricle as in chronic mitral regurgitation. An increase in diastolic volume in any of these examples would extend the curve upward and to the right.

determination (8). Although the data were not obtained simultaneously, the concept of differing *PV* relationships was introduced. The fluid-filled catheters employed, however, are subject to under- and overdamping, and diastolic pressures may be difficult to measure, especially in early diastole.

A few years ago, the ratio end-diastolic volume/end-diastolic pressure $V_{\text{LVED}}/P_{\text{LVED}}$ was taken as an index of the compliance of the left ventricle (9). Although this measurement provided some insights, it gave the relationship for only one point in the cardiac cycle and did not reflect the changing slope in the *PV* relationship during diastole. Others measured the changes in volume and pressure during atrial contraction to obtain two points from the *PV* relationship (10), allowing an estimate of compliance. Again, however, only a small segment of the *PV* curve is sampled in this sort of calculation.

In order to determine compliance ($\Delta V/\Delta P$), some investigators defined ΔP as the change from the lowest ventricular pressure during diastole to P_{LVED}, and ΔV as the stroke volume, obtained by angiographic or other means (11, 12). This approach uses information throughout diastole, but does not provide insight about the exact part of the ventricular *PV* curve at which filling is occurring. Although the data can be corrected for a given pressure (7) and a linear relationship obtained, it is still important to consider both the position of the *PV* data and the derivatives calculated from them.

The limitation in physiological application of compliance or stiffness calculations is important to recognize. These terms describe the slope of a curvilinear relationship but do not of themselves indicate where on the ventricular *PV* curve the measurements have been made.

Grossman and associates (13) utilized the echocardiogram for volume measurements and obtained simultaneous pressures to construct *PV* relationships. This approach described the *PV* curve throughout filling, without the technical problems of simultaneous, continuous angiocardiographic and pressure measurements during diastole. However, there are some major uncertainties about the precision of volume determinations by echocardiography. With the development of high-fidelity catheter-tip manometers, simultaneous pressure and volume data of high quality can now be obtained (14). Although such studies still have stringent technical requirements and necessitate left ventricular catheterization, they are becoming more frequent.

In evaluating reports of myocardial stiffness and compliance, one should look carefully at the methods by which the pressure and volume data were obtained and analyzed. Indices that sample diastole at only one point have limitations, as do stiffness and compliance when considered apart from specific volumes or pressures. Despite limitations in many studies, however, abnormalities found in the *PV* relationship have led to a greatly broadened understanding of cardiac pathophysiology.

CLINICAL IMPLICATIONS

In the normal ventricle, filling during diastole results in little change in the left ventricular pressure because of a relatively flat *PV* curve (Figure 1). Thus, ventricular filling and stroke volume can vary over a substantial range, yet P_{LVED} and pulmonary capillary pressure remain low.

Rightward Shift of PV Curve

In patients with chronic aortic or mitral regurgitation or congestive cardiomyopathy, the left ventricle is markedly dilated, yet may accommodate its filling volume with only a modest change in pressure. This is achieved by a shift of the *PV* relationship to the right (Figure 1) (8, 15). In the case of the regurgitant valve, the filling volume (total left ventricular stroke volume) is considerably greater than normal. Thus, an enlarged ventricle can accept a large inflow volume and deliver a normal forward stroke volume without marked elevation of P_{LVED} and pulmonary capillary pressure. Unless the *PV* curve were to shift to the right, the large V_{LVED} would necessitate filling pressure near pulmonary edema levels. Although this may ultimately occur, the diastolic volume is by then very large.

Leftward Shift of PV Curve

In sharp contrast, other states produce a shift of the *PV* curve such that higher pressures result at a given volume. The pathophysiology of pulmonary congestion and edema is better understood if this change in the *PV* relationship is appreciated.

Experimental data begin to explain some of the clinical changes observed. Ischemia or hypoxia shifts the pressure-volume curve to the left; for a given volume, the pressure is higher, and small changes in volume can result in markedly elevated pressures (Figure 1). In isolated muscle and intact canine hearts, hypoxia and tachycardia due to pacing produce incomplete muscle relaxation (16). Peak negative dP/dt, an index of myocardial relaxation in these experiments, fell with hypoxia. At a constant volume, the diastolic pressure increased, which indicated a leftward shift in the *PV* curve and probably increased stiffness.

Hood and colleagues (17) showed that compliance was decreased within three to five days after acute infarction in canine hearts. Pressure-length measurements in the infarcted segment compared with the adjacent normal tissue demonstrated that stiffness was increased only in the infarcted muscle.

Analogous results have been found in patients with ischemic heart disease. Abnormally high P_{LVED} was reported in patients who did not have high V_{LVED} (10), and decreased compliance was noted in patients studied within one year of a myocardial infarction (12). McLaurin and associates (18) found a fall in peak negative dp/dt, an increase in P_{LVED}, and a decrease in V_{ED} in patients with coronary artery disease. Using angiographic data and diastolic pressures, two groups reported decreased compliance (increased stiffness) with atrial pacing in patients with chronic ischemic heart disease, presumably due to induced ischemia (19, 20).

The increased ventricular pressure at a given volume is an important finding in patients with coronary artery disease, whether present chronically or associated with acute ischemia. Further increments in volume will cause the pressure to go higher yet, i.e. less reserve for an increase in diastolic volume is present. Thus, such patients may be in increased jeopardy for pulmonary edema (21, 22), changing from normal to high pulmonary capillary pressures with only modest changes in V_{LVED}.

In aortic stenosis, the pressure-volume curve is also shifted to the left; for normal volumes there is an elevated pressure (13, 23). In patients with hypertrophic subaortic stenosis, the V_{LVED} may even be low with elevated pressures, indicating a curve even further to the left (Figure 1) (6, 24). These changes in volume stiffness in aortic stenosis have been attributed to the hypertrophy of the ventricular wall; the increased thickness of the myocardium may prevent normal expansion during physiologic diastolic pressures, therefore higher pressures are necessary to achieve a normal end-diastolic volume and cardiac output. The muscle itself could have normal physical characteristics per unit mass.

The importance of a leftward shift of the *PV* curve is evident when an adaptation to changing body needs for blood flow is considered. When the cardiac output is sensed as inadequate, there are mechanisms available to increase cardiac filling and diastolic size by venous contraction and/or an increased blood volume. With the normal *PV* curve, a significant increase in diastolic volume can occur without approaching a diastolic pressure near pulmonary edema level. With a steeper *PV* curve that is also shifted toward the left, any modest increase in volume is associated with a more marked rise in pressure, and development of symptoms (21, 22).

ATRIAL CONTRACTION AND VENTRICULAR STIFFNESS

Figure 2 displays a left ventricular diastolic pressure tracing as it appears normally and in several diseases. The ventricular diastolic pressure usually reported is at the end of all filling (P_{LVED}) measured after the *a* wave. The *a* wave results from atrial systole and its size correlates with left ventricular stiffness, in turn related to the force of atrial contraction. In disorders with increased stiffness and a leftward shift of the *PV* curve, such as aortic stenosis or ischemic heart disease, the *a* wave may be very prominent (Figure 2 *C*). In these patients there may be a marked difference between left ventricular end-diastolic pressure and mean diastolic pressure. This difference reflects an important physiological compensation. Mean diastolic LV pressure (and pulmonary capillary pressure) will be lower if a given V_{LVED} and P_{LVED} are produced suddenly by an increased filling rate at end-diastole. This can be achieved if passive filling early in diastole is relatively slow, and if atrial contraction transmits a large proportion of filling in the last brief part of diastole. This occurs in aortic stenosis (23) and probably also in other states with equivalent left ventricular hypertrophy. Similar importance of atrial contraction has been described in coronary artery disease (25).

If this atrial booster-pump action were lost, as with atrial fibrillation, cardiac filling would fall and/or mean LV diastolic (and pulmonary capillary) pressure would rise in these disease states. A low cardiac output or pulmonary edema can be seen in that circumstance. In contrast, the changes in ventricular filling and pressure with loss of atrial contraction in a normal person are trivial.

The diastolic pressure tracing in chronic congestive cardiomyopathy is shown in Figure 2*B*. The initial LV diastolic pressure is elevated, and there is little difference between the mean diastolic and end-diastolic pressures. This pressure form is more likely to be found in a ventricle with a rightward shift of the *PV* curve. Pulmonary capillary pressure will be higher at a given P_{LVED}, in contrast to other disorders previously mentioned, such as aortic stenosis (Figure 2*C*).

Figure 2*D* shows left ventricular diastolic pressure in a patient with constrictive pericarditis. In this example mechanical restriction produces a steep *PV* curve. By mid-diastole, pressure markedly rises and plateaus to form the classical "square-root sign" for constrictive pericarditis. A similar pressure tracing also occurs in restrictive cardiomyopathies, in which the myocardial stiffness is altered.

EFFECTS OF TREATMENT ON LEFT VENTRICULAR STIFFNESS

The prospect of changing LV stiffness by shifting the *PV* curve is a new concept in cardiovascular therapeutics. There is not much information available, but two prototype studies can be cited. Miller and associates (26) reported LV compliance changed toward normal after coronary artery bypass graft surgery for ischemic heart disease.

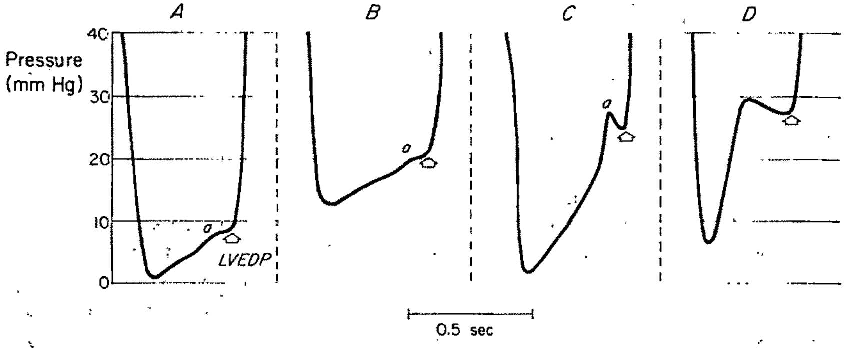


Figure 2 Diastolic left ventricular pressure tracings. End-diastolic pressure is shown by the arrow and the *a* wave is indicated. *A*. A normal individual with a P_{LVED} of 8 mm Hg and mean diastolic pressure of 5 mm Hg. *B*. Chronic congestive cardiomyopathy with elevated pressure at beginning and end of diastole. *C*. Chronic ischemic heart disease with very prominent *a* wave, and relatively low early diastolic pressure. *D*. Constrictive pericarditis exhibiting a classical "square-root sign" curve. There is an early diastolic dip, then the pressure quickly rises in mid-diastole and plateaus to end-diastole, without definite *a* wave.

Brodie et al (14) expanded our knowledge of the effects of nitroprusside in recent, elegant clinical experiments (Figure 3). A consistent, downward shift of the *PV* relationship during drug infusion was documented. This is an exciting area for future research. Even if the inotropic state of a given diseased heart could not be changed, a decrease in congestive symptoms should result, if the *PV* curve could be shifted by pharmacological means.

COMMENTS

The *PV* relationship can be described in a far more detailed, complex manner than herein (27). For clinical insight, however, it is important not to lose sight of the broad concepts, while research studies further evaluate the details of the LV properties. Indeed, some papers have simply described *PV* relationships as such, without calculations of stiffness or compliance (14, 20). Recently, because of problems of concept and measurement, the recommendation has been made to evaluate only V_{LVED} and P_{LVED} (28). This approach would not provide adequate information for clinical research, but the paper cited shows that calculations of stiffness and compliance currently used are not sufficient to discriminate between groups of patients.

Much of the work with *PV* curves assumes that the ventricle acts as a purely elastic chamber, i.e. the rate of change of volume is not important in determining the associated pressure. Put another way, the same *PV* curve would result whether the ventricle filled rapidly or slowly. We have wondered for a number of years, however, whether there might be a viscous component, which would also

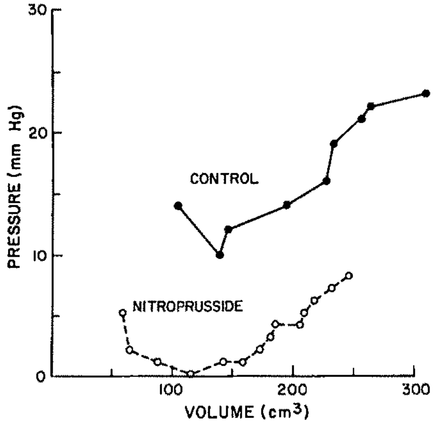


Figure 3 Reproduced from Brodie et al (14), this shows the effect of nitroprusside infusion on the diastolic pressure-volume curve from a patient with chronic mitral regurgitation. There is a substantial downward displacement during the infusion of nitroprusside, with the ventricular pressure being lower for any given volume. It is noteworthy that the shape of the curve (or slope of the tangent) is not significantly altered by the nitroprusside.

alter the *PV* curve (29). Recent work again raises this issue and may make the instantaneous rate of filling an influence on LV stiffness to be considered in the future (30, 31).

CONCLUSIONS

Left ventricular stiffness is defined by the relationship between pressure and volume and can be altered by a variety of disorders. In disorders with high V_{LVED}, such as mitral regurgitation, protection of pulmonary capillaries from high pressure is provided by a *PV* shift. In contrast, increased stiffness is produced by ventricular hypertrophy or ischemic heart disease, which puts the patient at increased risk for pulmonary capillary hypertension. Changing the ventricular *PV* relationship by therapy is an exciting prospect.

Literature Cited

1. Eichna, L. W. 1960. Circulatory congestion and heart failure. *Circulation* 22:864–86
2. Braunwald, E., Ross, J. Jr. 1963. The ventricular end-diastolic pressure. Appraisal of its value in the recognition of heart failure in man. *Am. J. Med.* 34:147
3. Bristow, J. D., Crislip, R. L., Farrehi, C., Harris, W. E., Lewis, R. P., Sutherland, D. W., Griswold, H. E. 1964. Left ventricular volume measurements in man by thermodilution. *J. Clin. Invest.* 43: 1015–24
4. Grossman, W., McLaurin, L. P., Moos, S. P., Stefadouros, M., Young, D. T. 1974. Wall thickness and diastolic properties of the left ventricle. *Circulation* 49:129–35
5. Covell, J. W., Ross, J. Jr. 1973. Nature and significance of alterations in myocardial compliance. *Am. J. Cardiol.* 32:449–55
6. Gaasch, W. H., Levine, H. J., Quinones, M. A., Alexander, J. K. 1976. Left ventricular compliance: Mechanisms and clinical implications. *Am. J. Cardiol.* 38:645–53
7. Diamond, G., Forrester, J. S. 1972. Effect of coronary artery disease and acute myocardial infarction on left ventricular compliance in man. *Circulation* 40:11–19
8. Dodge, H. T., Hay, R. E., Sandler, H. 1962. Pressure-volume characteristics of the diastolic left ventricle of man with heart disease. *Am. Heart J.* 64:503–11
9. Dwyer, E. M. Jr. 1970. Left ventricular pressure-volume alterations and regional disorders of contraction during myocardial ischemia induced by atrial pacing. *Circulation* 42:1111–22
10. Bristow, J. D., Van Zee, B. E., Judkins, M. P. 1970. Systolic and diastolic abnormalities of the left ventricle in coronary artery disease: Studies in patients with little or no enlargement of ventricular volume. *Circulation* 42:219–28
11. Gotsman, M. S., Lewis, B. S. 1974. Left ventricular volumes and compliance in hypertrophic cardiomyopathy. *Chest* 66:5, 498–505
12. Smith, M., Russell, R. O. Jr., Moraski, R. E., Feild, B. J., Rackley, C. E. 1974. Left ventricular *A* wave amplitude in patients after myocardial infarction. *Am. J. Cardiol.* 33:370–7
13. Grossman, W., Stefadouros, M. A., McLaurin, L. P., Rolett, E. L., Young, D. T. 1973. Quantitative assessment of left ventricular diastolic stiffness in man. *Circulation* 47:567–74
14. Brodie, B. R., Grossman, W., Mann, T., McLaurin, L. P. 1977. Effects of sodium nitroprusside on left ventricular diastolic pressure-volume relations. *J. Clin. Invest.* 59:59–68
15. Gault, J. H., Covell, J. W., Braunwald, E., Ross, J. Jr. 1970. Left ventricular performance following correction of free aortic regurgitation. *Circulation* 42:773–85
16. Weisfeldt, M. L., Armstrong, P.,

Scully, H. E., Sanders, C. A., Daggett, W. M. 1974. Incomplete relaxation between beats after myocardial hypoxia and ischemia. *J. Clin. Invest.* 53:1626–36
17. Hood, W. B. Jr., Bianco, J. A., Kumar, R., Whiting, R. B. 1970. Experimental myocardial infarction IV. Reduction of left ventricular compliance in the healing phase. *J. Clin. Invest.* 49:1316–23
18. McLaurin, L. P., Rolett, E. L., Grossman, W. 1973. Impaired left ventricular relaxation during pacing-induced ischemia. *Am. J. Cardiol.* 32:751–57
19. Barry, W. H., Brooker, J. Z., Alderman, E. L., Harrison, D. C. 1964. Changes in diastolic stiffness and tone of the left ventricle during angina pectoris. *Circulation* 49:255–63
20. Mann, T., Brodie, B. R., Grossman, W., McLaurin, L. P. 1977. Effect of angina on left ventricular diastolic pressure-volume relationship. *Circulation* 55:761–67
21. Dodek, A., Kassebaum, D. G., Bristow, J. D. 1972. Pulmonary edema in coronary artery disease without cardiomegaly: A paradox of the stiff heart. *N. Engl. J. Med.* 286:1347–50
22. Dodek, A., Kassebaum, D. G., Bristow, J. D. 1973. Pulmonary edema without cardiomegaly: Ischemic cardiomyopathy and the small stiff heart. *Am. Heart J.* 85:281–84
23. Stott, D. K., Marpole, D. G. F., Bristow, J. D., Kloster, F. E., Griswold, J. E. 1970. The role of left atrial transport in aortic and mitral stenosis. *Circulation* 41:1031–41
24. Lewis, R. P., Bristow, J. D., Farrehi, C., Kloster, F. E., Griswold, H. E. 1965. Idiopathic left ventricular hypertrophy: A hemodynamic reappraisal. *Am. J. Med.* 38:842–52
25. Rahimtoola, S. H., Ehsani, A., Sinno, M. Z., Loeb, H. S., Rosen, K. M., Gunnar, R. M. 1975. Left atrial transport function in myocardial infarction: importance of its booster pump function. *Am. J. Med.* 59:686–94
26. Miller, R. R., DeMaria, A. N., Amsterdam, E. A., Mailander, M. M., Zelis, R., Lurie, A. J., Mason, D. T. 1975. Improvement of reduced left ventricular diastolic compliance in ischemic heart disease after successful coronary artery bypass surgery. *Am. J. Cardiol.* 35:11–16
27. Mirsky, I., Cohn, P. F., Levine, J. A., Gorlin, R., Herman, M. V., Kreulen, T. H., Sonnenblick, E. H. 1974. Assessment of left ventricular stiffness in primary myocardial disease and coronary artery disease. *Circulation* 50:128–36
28. Glantz, S. F. 1976. Ventricular pressure volume curve indices change with end-diastolic pressure. *Circ. Res.* 39:772–78
29. Bristow, J. D., Kloster, F. E., Farrehi, C., Brodeur, M. T. H., Lewis, R. P., Griswold, H. E. 1966. The effects of supine exercise on left ventricular volume in heart disease. *Am. Heart J.* 71:319–29
30. Gaasch, W. H., Cole, J. S., Quinones, M. A., Alexander, J. K. 1975. Dynamic determinants of left ventricular diastolic pressure-volume relations in man. *Circulation* 51:317–24
31. Rankin, J. S., Arentzen, C. E., McHale, P. A., Ling, D., Anderson, R. W. 1977. Viscoelastic properties of the diastolic left ventricle in the conscious dog. *Circ. Res.* 41:37–45

Ann. Rev. Med. 1978. 29:485–93
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CARDIOGENIC HYPERTENSION[1] ◆7293

Harriet P. Dustan, M.D. and Robert C. Tarazi, M.D.
Research Division, Cleveland Clinic Foundation, Cleveland, Ohio 44106

INTRODUCTION

Since the heart supplies the energy for the circulation of blood, it is an important determinant of arterial pressure. For hypotension, this role is clearly seen in cardiogenic shock. However, in the case of hypertension, the role of the heart is neither so obvious nor so well proven, and there are no widely accepted cardiogenic hypertensions. Probably this is because vascular resistance is also a determinant of pressure, and it is almost always deranged in hypertension (1). Thus, although cardiac output is elevated in several clinical hypertensions, so also is peripheral resistance, and it is difficult to determine the contribution of each factor.

There is much current interest in the role of the heart in the initiation and/or maintenance of clinical and experimental forms of hypertension, and much of this is focused on the importance of elevated cardiac output (2). There are two mechanisms for increasing output: increased central blood volume and enhanced myocardial contractility. The former reflects not only the total blood volume but also the capacity of reservoir vessels below heart level; even without volume expansion it may be elevated because of venoconstriction and shunting of blood centrally (2). This venoconstriction is thought to be of neural origin (3). Enhanced myocardial contractility can initiate a rise in output by increasing the ejection fraction, and possibly also by drawing on part of the central blood volume (4). However, that source is limited and cardiac output can remain elevated only if concomitant adjustments in the peripheral circulation allow a rapid venous return (5). The enhanced myocardial contractility is usually a reflection of increased cardioadrenergic drive, although serum electrolytes and humoral factors may play an additional role in some instances (6–8). Thus, cardiogenic hypertension cannot be considered without considering the role of the autonomic nervous system, particularly its sympathetic component.

Elevated cardiac output is thought to initiate some experimental hypertension. In the early stages of the renal form, the rise in arterial pressure is due to

---

[1] Studies referred to in this review were supported in part by grants from the National Heart, Lung, and Blood Institute (NIH-HL 6835), John A. Hartford Foundation, and the Whitaker Foundation.

0066-4219/78/0401-0485$01.00

increased cardiac output, since vascular resistance is normal or slightly decreased (9, 10). A bit later, resistance rises and cardiac output falls to control levels, so that hypertension is then sustained by an elevated vascular resistance. This hemodynamic sequence is ascribed to "total body autoregulation" (11) and is usually explained as a compensatory arteriolar myogenic response to increased flow (12). Whether or not this sequence is essential to the development and maintenance of all experimental hypertensions is a contemporary debate (13). The same debate extends to clinical hypertensions (2, 13, 14). Many types are associated with an elevated output (1, 2, 15, 16), but the causal role of an increased cardiac output even in these types of the disease is far from proven (1). At this point it is important to emphasize that it is not possible to conclude from the presence of a high output in a hypertensive state that the rise in arterial pressure is "cardiogenic" (17).

All of the foregoing considerations have dealt with cardiac hemodynamic factors as determinants of arterial pressure, yet the heart may well have another role in hypertension. Emerging evidence indicates that it can be a source of powerful pressor reflexes (18–20). Whether these have an ongoing function in chronic hypertension is yet to be determined, although it seems likely that they may be responsible for the rises in arterial pressure that accompany angina and myocardial infarction (18–22) or that follow coronary revascularization procedures (23, 23a, 24).

Thus the heart can be actively related to hypertension in different ways; the purpose of the following discussion is to evaluate the evidence for this relationship. To that purpose however, a clarification of the term used, cardiogenic hypertension, is essential. It refers to a hypertension in which the heart plays a major role either by initiating an increase in output or as the source of pressor reflexes. To the latter instance, the term cardiogenic is strictly applicable. In the former situation, however, the term is acceptable only as an approximation, since the mechanism increasing cardiac performance may have its origin outside the heart.

HYPERTENSION WITH INCREASED CARDIAC OUTPUT

Clinical Studies

It has only been within the past ten years that the hemodynamics of hypertension have been extensively surveyed (1, 2, 15). These studies revealed a marked unhomogeneity between types of hypertension and often within types. We now accept the fact that some hypertensions tend to have increased cardiac output of mild to moderate degree (15, 25). These are borderline hypertension, renovascular hypertension, primary aldosteronism, the hypertension of end-stage renal disease, and a type of established essential hypertension with a hyperkinetic circulation. Although this list covers some patients with essential hypertension and most of those with secondary types of hypertensions, it concerns a relatively small proportion of the hypertensive population, since most patients have established essential hypertension with normal or decreased cardiac output (1).

Description of these high output types does not define the role of increased flow in the elevated pressure, but that is not an easy task. When propranolol first became available it was anticipated that its use would define those patients with cardiogenic hypertension in that the resultant decrease in cardiac output would be followed promptly by a fall in arterial pressure. This proved not to be the case, because although propranolol administered intravenously decreased cardiac output rapidly, vascular resistance rose and arterial pressure did not fall. Only later, and on chronic therapy, did arterial pressure fall, and even then those patients who responded were not necessarily those with pretreatment elevated cardiac output (26; Figure 1).

Another way to examine the data is to correlate arterial pressure with cardiac output. In renovascular hypertension these two variables correlated positively, which suggested that the increased output contributed to the hypertension (27). In primary aldosteronism, however, pressure and output were inversely related (7). In end-stage kidney disease, the high cardiac output is due to the anemia and seems to be insignificant in regard to the hypertension, since normotensives have equally high output; the feature distinguishing the two groups of uremic patients is peripheral vascular resistance, which is elevated in those with elevated pressure and low in those with normal pressure (28).

BETA-ADRENERGIC BLOCKADE, CARDIAC OUTPUT, ARTERIAL PRESSURE

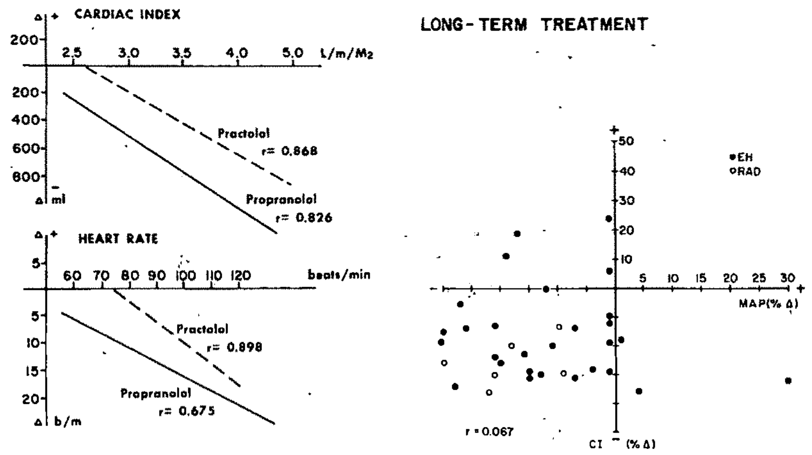


Figure 1 Whereas the immediate cardiac output response to propranolol is related to its resting level and could be taken as an index of cardioadrenergic activity (*left panel*), the arterial pressure response to long-term therapy is not correlated with changes in output (*right panel*) and cannot be used to define "cardiogenic hypertension." Based on data from Tarazi & Dustan (26) (Reproduced by permission).

The absence of a relationship between cardiac output and arterial pressure in chronic hypertension might be ascribed to the development of structural changes (hypertrophy) in the heart (28–30) but specially in the arterioles. The former limits compliance of the ventricle, while the latter helps maintain the high peripheral resistance by an increased arteriolar wall/lumen ratio (31). There is evidence for and against the importance of vascular hypertrophy. Supporting evidence is the slow fall in arterial pressure and vascular resistance that occurs during chronic propranolol therapy (26). After six months of treatment, we found vascular resistance still elevated above control values, and only after a year of therapy was any real decrease found. These slow changes could reflect vascular remodeling. The evidence against the cardinal importance of arteriolar thickening is the response to vasodilator drugs, such as diazoxide and minoxidil. These drugs lower pressure rapidly by arteriolar dilatation without affecting veins (32); as long as the drugs are given they are effectively antihypertensive unless some compensatory, nullifying response supervenes (33). Similarly, renovascular hypertension abates rapidly with reduction in total peripheral resistance (TPR) following successful vascular repair or nephrectomy (27). These observations indicate that luminal narrowing is not an immutable feature of chronic hypertension.

Experimental Hypertensions

The doubts raised by clinical studies regarding the importance of an elevated output for the maintenance of hypertension leave open the possibility that an increase in output is important in initiating the rise in arterial pressure. This question is best addressed in well-defined experimental models that allow prospective sequential studies.

As indicated above, many experimental renal hypertensions begin with an increase in cardiac output followed by an increase in peripheral resistance. How important is the elevated output for this sequence? Some investigators consider it to be causal (11). However, the possibility that it may be an incidental transient has not been investigated in these forms, although it received considerable attention in other forms and this evidence is worth noting. Studies have been performed in different models; some (obviously salt- and water dependent) in which the rise in output was related to expanded extracellular fluid volume (13, 34), and others where neurogenic factors and increased cardiac performance were thought to predominate (35, 36). These latter include spontaneous hypertension in rats and chronic stellate stimulation in dogs.

Young spontaneous-hypertensive (SH) rats (Okamoto strain) have increased cardiac output; as they age, this remits. During most of the natural course of their hypertension, elevated arterial pressure reflects an increased vascular resistance (35). However, if pregnant SH rats are treated with propranolol and their offspring maintained continuously on the drug from birth, these young SH rats will not have a phase of increased cardiac output, yet subsequent development of hypertension will be uninhibited (35).

Sustained hypertension can be produced in dogs by electrical stimulation of the left stellate ganglion. After an initial phase of high output, systemic flow returns to normal within one day, vascular resistance rises, and subsequently, hypertension reflects the increased resistance (36). Here again however, the initial rise in output is not necessary for development of hypertension. Its prevention by propranolol and a constant-flow pump has no influence at all on the time of onset or the degree of subsequent hypertension (4, 36).

Metyrapone-induced hypertension in dogs is a result of ACTH overdrive and enhanced desoxycorticosterone production. Sequential studies reveal three hemodynamic patterns (13, 34). In one, the classic autoregulatory type, an early increase in cardiac output and pressure is followed by a rise in resistance and normalization of output, with hypertension maintained by high TPR. Another type has high cardiac output and high pressure with normal resistance as long as the drug is given. The third has increased resistance as a cause of hypertension from the very beginning without a rise in output. Finally, renal hypertension in the rabbit (in contrast to dogs) develops without an early transient period of increased blood flow; the pressure rise occurs from the beginning because of raised resistance (37).

In summary, current information does not support the conclusion that an elevated cardiac output in various types of hypertension is the cause of elevated arterial pressure, or that it is a prerequisite in its hemodynamic evolution. However, this negative conclusion does not mean that a rise in output has no significance at all. It might not be the sole or even the most important factor, but it may influence clinical characteristics and play an additive role in some conditions.

Mechanisms of Increased Cardiac Output

As indicated earlier, there are two ways that cardiac output can be elevated; (*a*) increase in central blood volume because of either hypervolemia or central redistribution of blood, and (*b*) enhanced myocardial performance with increased ejection fraction. Indices defining both functions can be measured by standard (2) or even by noninvasive techniques (38). The ratio of cardiopulmonary volume (CPV) to total blood volume (TBV) is an estimate of the distribution of blood between the peripheral and central circulations (2). A significant increase indicates a redistribution of blood into the central circulation because of diminished capacity of reservoir vessels (venoconstriction) below heart level, while a significant decrease indicates increased capacity, presumably due to venodilatation. Both cardiac output and stroke volume are significantly correlated with CPV, but the correlation is closer with stroke volume (SV) than with cardiac output (CO) (2, 39). A significant increase of the SV/CPV ratio suggests enhanced myocardial performance, while a significant decrease suggests impaired cardiac performance (2).

The value of this approach was demonstrated in a study of two groups of hypertensives with elevated cardiac index, renovascular hypertensives and patients with borderline essential hypertension. Those with renovascular hypertension had a significant increase in CPV/TBV but CO/CPV was within the normal range. In contrast, the borderline hypertensives had a normal CPV/TBV ratio but increased CO/CPV, which indicates enhanced cardiac performance as the cause of their increased cardiac output (Figure 2).

A noninvasive and yet more direct approach to myocardial performance than determination of cardiac output is the recording of systolic time intervals. Careful classification of hypertensive patients can reveal significant differences that are hidden by consolidating in one group all patients with elevated arterial pressure (40). Equally important is the consideration of pressure levels in the evaluation of alterations in ejection time, pre-ejection period (PEP) or isovolumic contraction (IVCT) (40, 41). Thus, we (40) found that borderline hypertensives as well as some patients with severe, long-standing hypertension associated with marked fluctuations in blood pressure (25), had a significantly increased ratio of diastolic blood pressure (DBP) to IVCT in contrast to a normal ratio in patients with fixed essential hypertension [DBP/IVCT is a close correlate of maximum left ventricular (LV) d*p*/d*t* (41)].

Alterations of these indices by propranolol provide a good index of adrenergic drive to the heart (40, 42). Differences in DBP/IVCT among hypertensive patients were abolished by propranolol, which indicates the wide spectrum in

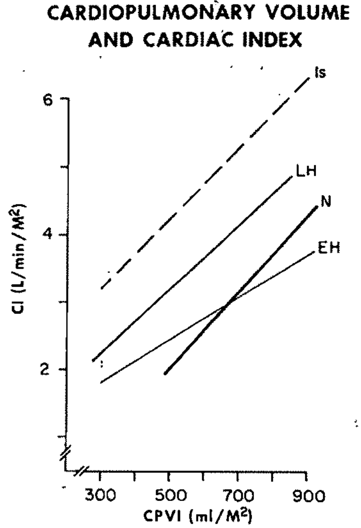


Figure 2 Regression lines defining the correlation between cardiac index and cardiopulmonary volume in 9 normotensive volunteers (N), 16 essential hypertensives (EH), and 11 patients with borderline (labile) hypertension (LH); IS defines the correlation obtained in 20 hypertensive patients given isoproterenol infusion (0.03 μg/kg per min). The shift of the line up to the left in LH as compared with EH was similar to that product by isoproterenol. From Tarazi et al (2), and reproduced by permission of the American Heart Association Inc.

cardioadrenergic drive in hypertension. The clinical importance of such investigations in special instances is the determination of the degree of dependence of cardiac function on sympathetic drive.

PRESSOR REFLEXES ARISING FROM THE HEART

For a long time it has been known that anginal attacks may be preceded by a rise in arterial pressure and that myocardial infarction may be accompanied by transient hypertension (21, 22). More recently, other transient pressor events connected with the heart have been recognized. Patients undergoing surgical replacement of the aortic valve may develop hypertension postoperatively (43), and this can contribute to graft failure when homografts are used. Also paroxysmal hypertension is now recognized as a frequent occurrence in the early hours following an aortic–coronary arterial bypass graft (23). It can be controlled by either sodium nitroprusside given intravenously or stellate ganglion infiltration with procaine (23, 24).

These hypertensions may be an expression of the powerful pressor reflex role of the heart. This can be triggered from either distension of coronary vessels (18, 19), myocardial ischemia (44), or specific chemoreceptors (20, 45). Small doses of 5-hydroxytryptamine (serotonin) given into the left coronary artery can evoke a rapid rise in arterial pressure in the dog (45). This does not occur when serotonin is injected into the distal left coronary artery; instead the depressor reflex of Bezold-Jarisch is evoked causing bradycardia and hypotension (20). James et al (20) showed that this cardiogenic reflex was initiated by a small chemoreceptor close to the main left coronary artery and supplied by it and that its afferents were vagal. The hypertension produced is strikingly severe; it lasts about ten minutes and is associated with vagal and sympathetic efferent effects in both sinus node and A-V junction causing changes in cardiac rate and rhythm.

Other reflexes from the heart originate from distension of coronary vessels; those mediated by vagal afferents lead to hypotension, those mediated by sympathetic afferents raise arterial pressure (18, 19). Of particular relevance in this context is the recent report that post-myocardial-revascularization hypertension and increased TPR can often be controlled by unilateral stellate block (24).

SUMMARY

The importance of the heart as a source of pressor reflexes is just beginning to be investigated. Clinical experience suggests that it will be shown to have great significance, not only for angina, myocardial infarction, and cardiac surgical procedures, but possibly also in chronic sustained hypertension.

Literature Cited

1. Dustan, H. P., Tarazi, R. C., Hinshaw, L. B. 1976. Mechanisms controlling arterial pressure. In *Pathophysiology Arterial Regulatory*

Mechanisms in Diseases, ed. E. D. Frohlich, pp. 48–82. Philadelphia: Lippincott. 902 pp. 2nd ed.

2. Tarazi, R. C., Ibrahim, M. M., Dustan, H. P., Ferrario, C. M. 1974. Cardiac factors in hypertension. *Circ. Res* 34 (Suppl. 1):213–21
3. Yu, P. N. 1969. *Pulmonary Blood Volume in Health and Disease*, pp. 122–88. Philadelphia: Lea & Febiger
4. Liard, J. F., Tarazi, R. C., Ferrario, C. M. 1975. Hemodynamic effects of stellate ganglion stimulation in conscious dogs. In *Nervous System in Arterial Hypertension*, eds., S. Julius, M. Esler, pp. 151–61. Springfield, Ill.: Thomas
5. Guyton, A. C., Jones, C. E., Coleman, T. G. 1973. Introduction to the regulation of cardiac output. In *Circulatory Physiology: Cardiac Output and Its Regulation*, pp. 137–46. Philadelphia: Saunders
6. Haddy, F. J., Scott, J. B., Emerson, T. E., Overbeck, H. W., Daugherty, R. M. Jr. 1969. Effects of generalized changes in plasma electrolyte concentration and osmolarity in blood pressure in the anesthetized dog. *Circ. Res.* 24 & 25 (Suppl. 1):I59–I74
7. Tarazi, R. C., Ibrahim, M. M., Bravo, E. L., Dustan, H. P. 1973. Hemodynamic characteristics of primary aldosteronism. *N. Engl. J. Med.* 289:1330–35
8. London, G. M., Safar, M. E., Weiss, Y. A., Corvol, P. L., Menard, J. E., Simon, A. C., Milliez, P. L. 1977. Relationship of plasma renin activity and aldosterone levels with hemodynamic function in essential hypertension. *Arch. Intern. Med.* 137:1042–47
9. Ledingham, J. M., Cohen, R. D. 1963. Role of the heart in the pathogenesis of renal hypertension. *Lancet* 2:979–81
10. Ferrario, C. M., Page, I. H., McCubbin, J. W. 1970. Increased cardiac output as a contributory factor in experimental renal hypertension in dogs. *Circ. Res* 27:799–810
11. Guyton, A. C., Coleman, T. G. 1969. Quantitative analysis of the pathophysiology of renal hypertension. *Circ. Res.* 24, 25 (Suppl. 1):1–19
12. Bayliss, W. M. 1902. On the local reactions of the arterial wall to changes of internal pressures. *J. Physiol.* 28:220–231
13. Tarazi, R. C. 1976. Hemodynamic role of extracellular fluid in hypertension. *Circ. Res.* 38 (Suppl. 2): 72–83
14. Onesti, G., Kim, K. E., Greco, J. A., Del Guercio E. T., Fernandes M., Schwartz C. 1975. Blood pressure regulation in end-stage renal disease and anephric man. *Circ. Res.* 36, 37 (Suppl. 1):145–52
15. Frohlich, E. D. 1977. Hemodynamics of hypertension. *Hypertension Physiopathology and Treatment*, eds. J. Genest, E. Koiw, O. Kuchel, pp. 15–48. New York: McGraw-Hill. 1208 pp.
16. Julius, S., Conway, J. 1968. Hemodynamic studies in patients with borderline blood pressure evaluation. *Circulation* 38:282–88
17. Tarazi, R. C., Ferrario, C. M., Dustan, H. P. 1977. The heart in hypertension. See Ref. 15, pp. 738–54
18. Brown, A. M. 1967. Excitation of afferent cardiac sympathetic nerve fibres during myocardial ischaemia. *J. Physiol. London* 190:35–53
19. Malliani, A., Brown, A. M. 1970. Reflexes arising from coronary receptors. *Brain Res.* 24:352–55
20. James, T. N., Isobe, J. H., Urthaler, F. 1975. Analyses of components in a cardiogenic hypertensive chemoreflex. *Circulation* 52:179–92
21. Horwitz, D. Sjoerdsma, A. 1965. *Proc. Counc. High Blood Pressure Res. Am. Heart Assoc.* 13:39–44
22. Levine, S. A., Ernstene, A. C. 1933. Observations on arterial blood pressure during attacks of angina pectoris. *Am. Heart J.* 8:323–26
23. Estafanous, F. G., Tarazi, R. C., Viljoen, J. F., El Tawil, M. Y. 1973. Systemic hypertension following myocardial revascularization *Am. Heart J.* 85:732–38

23a. Fouad, F. M., Estafanous, F. G., Tarazi, R. C. 1977. Hemodynamics of post-myocardial revascularization hypertension. *Am. J. Cardiol.* In press

24. Fouad, F. M., Estafanous, F. G., Tarazi, R. C. 1977. Post coronary-bypass hypertension: Evidence for its reflex origin. *Circulation* (Abstr.) 55, 56:Suppl. 3:30
25. Ibrahim, M. M., Tarazi, R. C., Dustan, H. P., Bravo, E. L., Gifford, R. W. Jr. 1975. Hyper-

kinetic heart in severe hypertension: A separate clinical hemodynamic entity. *Am. J. Cardiol.* 35:667–74

26. Tarazi, R. C., Dustan, H. P. 1972. Beta adrenergic blockade in hypertension: Practical and theoretical implications of long-term hemodynamic variations. *Am. J. Cardiol.* 29:633–40
27. Tarazi, R. C., Frohlich, E. D., Dustan, H. P. 1973. Contribution of cardiac output to renovascular hypertension in man: Relation to surgical treatment. *Am. J. Cardiol.* 31: 600–5
28. Neff, M. S., Kim, K. E., Persott, M., Onesti, G., Schwartz, C. 1971. Hemodynamics of uremic anemia. *Circulation* 43:876–83
29. Averill, D. B., Ferrario, C. M., Tarazi, R. C., Sen, S., Bajbus, R. 1976. Cardiac performance in rats with renal hypertension. *Circ. Res.* 38:280–88
30. Hallback, M., Isaksson, O., Noresson, E. 1975. Consequences of myocardial structural adaptation on left ventricular compliance and the Frank-Starling relationship in spontaneously hypertensive rats. *Acta Physiol. Scand. Suppl.* 424:1–12
31. Folkow, B. 1971. The hemodynamic consequences of adaptive structural changes of the resistance vessels in hypertension. *Clin. Sci.* 41:1–12
32. Tarazi, R. C., Dustan, H. P., Bravo, E. L., Niarchos, A. P. 1976. Vasodilating drugs: Contrasting haemodynamic effects. *Clin. Sci. Mol. Med.* 51 (Suppl. 3):575–78
33. Dustan, H. P., Tarazi, R. C., Bravo, E. L. 1976. False tolerance to antihypertensive drugs. In *Systemic Effects of Antihypertensive Agents*, ed. M. P. Sambhi, pp. 51–68. New York: Stratton Intercontinental Med. Book Corp. 596 pp.
34. Bravo, E. L., Tarazi, R. C., Dustan, H. P. 1977. Multifactorial analysis of chronic hypertension induced by electrolyte-active steroids in trained, unanesthetized dogs. *Circ. Res.* 40 (Suppl. 1):140–45
35. Pfeffer, M. A., Pfeffer, J. M., Weiss, A. K., Frohlich, E. D. 1977. Development of SHR hypertension and cardiac hypertrophy during prolonged beta blockade. *Am. J. Physiol.* 232:H639–H644
36. Liard, J. F., Tarazi, R. C., Ferrario, C. M., Manger, W. M. 1975. Hemodynamic and humoral characteristics of hypertension induced by prolonged stellate stimulation in conscious dogs. *Circ. Res.* 36:455–64
37. Fletcher, P. J., Korner, P. I., Argus, J. A., Oliver, J. R. 1976. Changes in cardiac output and total peripheral resistance during development of renal hypertension in the rabbit: Lack of conformity with the autoregulation theory. *Circ. Res.* 39:633–39
38. Falch, D., Norman, N. 1974. Evaluation of a computerized technique for determination of cardiac output and central circulation times using gamma camera and ^{113m} Indium. *Scand. J. Clin. Lab. Invest.* 34:207–14
39. Milnor, W. R., Jose, A. D., McGaff, C. J. 1960. Pulmonary vascular volume, resistance and compliance in man. *Circulation* 22:130–37
40. Ibrahim, M. M., Tarazi, R. C., Dustan, H. P., Bravo, E. L. 1974. Cardioadrenergic factor in essential hypertension. *Am. Heart J.* 88:724–32
41. Landry, A. B. Jr., Goodyer, A. V. N. 1965. Rate of rise of left ventricular pressure. Indirect measurement and physiologic significance. *Am. J. Cardiol.* 15:660–64
42. Harris, W. S., Schoenfeld, C. D., Weissler, A. M. 1967. Effects of adrenergic receptor activation and blockade on the systolic pre-ejection period, heart rate and arterial pressure in man. *J. Clin. Invest.* 46: 1704–14
43. Layton, C., Monro, J., Bridgen, W., McDonald, A., McDonald, L., Weaver, J. 1973. Systemic hypertension after homograft aortic valvar replacement. A cause of late homograft failure. *Lancet* 2:1343–47
44. Kent, K. M., Cooper, T. 1975. Editorial: Cardiovascular reflexes. *Circulation* 52:177–78
45. Eckstein, R. W., Shintani, F., Rowen, H. E. Jr., Shimomura, K., Ohya N. 1971. Identification of left coronary blood supply of aortic bodies in anesthetized dogs. *J. Appl. Physiol.* 30:488–92

Ann. Rev. Med. 1978. 29:495–507
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CELLULAR MECHANISMS IN ACID SECRETION

◆7294

Andrew H. Soll, M.D. and Morton I. Grossman, M.D.
School of Medicine, University of California, Los Angeles 90024, and Wadsworth Veterans Administration Hospital Center, Los Angeles, California 90073

The secretion of acid by the upper gastrointestinal tract first appeared among the lowest vertebrates on the phylogenetic tree and has since been retained by most species (1). The obvious physiological role for this acid is in the digestion of food by pepsin; however the importance of acid-peptic digestion in man appears quite limited in light of the normal nutritional status of persons with achlorhydria. The stomach may have retained the ability to secrete acid not primarily for digestion, but rather to sterilize the upper gastrointestinal tract. Although secretion of acid appears to be a dispensable function in man, acid does play a major role in the pathophysiology of peptic ulcer disease and of most of the bleeding disorders of the upper gastrointestinal tract. This review concerns our current, admittedly limited, knowledge of the cellular mechanisms of gastric acid secretion. It attempts to lay a groundwork for understanding certain aspects of acid-peptic disorders, particularly the pathophysiology and the mechanisms underlying the effectiveness of some of the therapeutic modalities used.

THE PHYSIOLOGY OF GASTRIC ACID SECRETION

The regulation of acid secretion ranks as one of the most complex physiological processes in the body. Previous teaching has been that the acid secretory response to a meal involves cephalic, gastric, and intestinal phases. Although this classification remains valid, when one is focusing on the cellular aspects of these processes it is easier to classify the pathways according to the final effectors at the cellular level (Table 1).

Four Pathways to the Parietal Cell

All bodily processes including gastric acid secretion, are regulated by chemical messengers, which originate in cells responsive to perturbations in the environment. They are delivered to target cells by one of three modes: neurocrine, endocrine, or paracrine. In response to a physiological stimulus such as a meal, the parietal cell is directly or indirectly stimulated by each of these pathways. In

0066-4219/78/0401-0495$01.00

Table 1 Major pathways for parietal cell stimulation

Type of transmission	Route	Chemical messenger
Neurocrine	Synaptic	Acetylcholine ? others
Endocrine	Blood	Gastrin ? others
Paracrine	Extra cellular	Histamine ? others
Luminal	Direct effects	Amino acids ? others

addition, recent studies indicate that luminal contents may also have a direct effect on the parietal cell. The relative importance of each of these pathways in determining the integrated response to a meal is not yet known. It is likely that these four pathways will not have the same relation under all physiological situations and in all species. In light of the fact that acid secretion is apparently a dispensable function, this redundancy of control is of interest.

NEUROCRINE Neuronal or neurocrine effects can directly activate the parietal cell to secrete acid by release of acetylcholine from postganglionic fibers. The neuronal path can be activated both by vagal efferent fibers and by local reflex arcs within the wall of the stomach (2, 3). Cephalic phase stimuli act solely through vagal efferents, whereas gastric distention can act both through vagovagal reflexes and also, somewhat less efficiently, through local reflexes. The effects mediated by both the vagal efferent fibers and local reflex arcs are inhibited by muscarinic cholinergic antagonists such as atropine. The view that the postganglionic neuronal pathway utilizes cholinergic transmitters is further supported by the observations in laboratory animals that exogenous cholinomimetic drugs such as carbachol produce the same maximal response as vagal stimulation, gastrin, or histamine. In man, however, there is only a tiny response to such drugs (4). These observations in man may reflect cholinergic inhibitory pathways or noncholinergic vagal efferent fibers. This difference highlights the wide variability between species in the regulation of gastric acid secretion.

The release of gastrin is also in part under neuronal control; each of the mechanisms just described for neuronal activation of the parietal cell can also cause neuronal release of gastrin (5). While it is clear that vagal stimulation, e.g. by sham feeding, causes release of gastrin in man and laboratory animals, administration of exogenous cholinomimetic drugs causes little or no release of gastrin, and atropine often enhances rather than inhibits gastrin release to many stimuli. This raises the question of whether the vagal release of gastrin is noncholinergic, or cholinergic but atropine resistant, or some other not yet dreamt of alternative.

ENDOCRINE The only hormone known to stimulate the parietal cell under physiological conditions is gastrin (5). Gastrin is released in response to vagal stimuli and antral distension, as was mentioned above, but it is mainly released by the direct luminal effects of food, particularly amino acids and partially digested protein (2, 5). It is likely that another hormone is involved in the intestinal phase of gastric secretion, but it has not yet been isolated (2, 6, 7).

PARACRINE Paracrine secretion refers to diffusion of a chemical messenger from its cell of origin across intercellular spaces to its target (8). Both the concept of paracrine secretion and its potential role in the secretion of acid remain hypothetical. There is, however, little doubt that histamine secreted locally from mucosal stores has a major effect on parietal cell function. Histamine appeared in the upper gastrointestinal tract at the same point in the phylogenetic tree that the ability to secrete acid was acquired (1). Histamine was shown to be a potent stimulant of acid secretion (9), an effect not inhibited by the classic antihistamines (H_1 antagonists), which suggests the existence of two classes of histamine receptors. However, the physiological role of histamine in gastric acid secretion was controversial until 1972. In that year, Black and co-workers (10) introduced burimamide, a competitive antagonist for the action of histamine on the H_2 receptors in stomach, atrium, and uterus. This agent and its two analogues, cimetidine (11) and metiamide, are specific antagonists of the H_2 effects of histamine. It is of great interest that these agents inhibit the secretory response not only to histamine but also to food, gastrin, and cholinergic stimuli (12). These findings established beyond question that histamine has important physiological effects on the parietal cell. However, our knowledge of histamine in the gastric mucosa remains quite limited. The histamine-containing cell has not been definitely identified, although there is some evidence that, in species other than rat, histamine is contained in a cell that is morphologically indistinguishable from a mast cell (13, 14). However, gastric histaminocytes are physiologically quite distinct from perivascular mast cells. Perivascular mast-cell histamine, but not gastric mucosal histamine, is released by Compound 48/80 and by antigen-antibody reactions (13). The mechanisms controlling the release of gastric histamine and the biochemical pathways accounting for the formation of gastric histamine stores remain unknown (15, 16). The rat is the only exception, for in this species histidine decarboxylase, the enzyme that decarboxylates histidine to form histamine, is present at high activity apparently in an enterochromaffin-like cell in the gastric mucosa (17). Furthermore, in rat this enzyme activity is induced by feeding and by gastrin, which thus supports the view that physiological stimuli induce the synthesis and therefore probably the release of histamine (16, 17). In other species, both the source of histamine and the factors controlling its synthesis and release remain unknown.

DIRECT Protein digests and amino acids stimulate acid secretion when instilled into isolated, denervated fundic pouches (2, 18) and thus appear to stimulate the parietal cell directly. Undigested protein, such as serum albumin, does not stimulate acid secretion (2). Amino acids also are potent stimulants when infused intravenously (19).

Other Regulators of Gastric Secretion

There are numerous endogenous agents that have been shown under experimental conditions to inhibit acid secretion. These endogenous inhibitory agents may act through paracrine, endocrine, and neurocrine pathways as described for the stimulants of acid secretion. Potential inhibitors that have been found in gastric mucosa and are thus candidates for paracrine inhibitors include somatostatin (8, 20, 21), the prostaglandins (22, 23), glucagon (21, 24), dopamine (25), and serotonin (26). Vasoactive intestinal peptide (VIP) has been localized to nerve terminals in gastric mucosa and has been proposed as a potential inhibitory neurotransmitter (21). A similar role as a neurotransmittor has also been proposed for dopamine (25). The presence of fat in the duodenum inhibits acid secretion by denervated fundic pouches, which suggests the presence of an inhibitory hormone; gastric inhibitory peptide (GIP) at present is the most likely candidate for this function (7, 27). Secretin inhibits acid secretion, but whether it operates within the physiological range is at present doubtful (2).

These agents have each been shown to have pharmacological effects that inhibit acid secretion; however the physiological significance of inhibition by any one of these agents is unknown and is extremely difficult to study, because, with the exception of secretin and GIP, they are secreted locally and generally undergo rapid degradation. Concentrations in blood and gastric juice, if measurable, are poor indicators of local tissue concentration. With the prostaglandins, however, their physiological role may be indirectly assessed by the effects of aspirin and the other nonsteroidal anti-inflammatory agents. The ulcerogenic properties of these anti-inflammatory drugs apparently do not depend solely upon breaking of the gastric barrier (28), and since these drugs inhibit prostaglandin synthesis (29), it is possible that their ulcerogenic properties result from loss of the normal cytoprotective and/or antisecretory functions of the prostaglandins.

A few other agents warrant consideration as important stimulants of acid secretion. Caffeine is a weak stimulant of acid secretion, causing 10% of the maximal pentagastrin stimulation in normal subjects (2). For reasons that have not been elucidated, caffeine is a more potent stimulant in duodenal ulcer patients, causing stimulation equal to 40% of the pentagastrin maximum. Since caffeine is a phosphodiesterase inhibitor, this effect could reflect enhanced accumulation of cyclic AMP in the parietal cell. Caffeine does not cause significant elevations of serum gastrin.

Alcohol stimulates acid secretion and gastrin release in the dog but not in man (2). The relation of alcohol to acute gastritis is probably due to "breaking" of the gastric mucosal barrier and not to direct effects on acid secretion.

The effect of calcium in gastric acid secretion depends upon the chronicity of the exposure and the associated disease states (2, 30). The acute administration of calcium either orally or intravenously stimulates acid secretion and serum gastrin in man but not in the dog. The total dose of calcium needed to produce an effect is much smaller by the oral than by the intravenous route; this then accounts for the significant stimulation of acid secretion by calcium-containing antacids. Chronic hypercalcemia such as that due to hyperparathyroidism is not associated with hypergastrinemia or hypersecretion unless the patient also has an associated gastrinoma, a pattern seen in multiple endocrine adenomatosis, Type I.

Interactions Between Secretagogues

The regulation of the secretion of acid would be adequately confusing were it to involve only the mechanisms just enumerated. However, to further complicate matters, gastric secretagogues do not act independently but rather show a marked interdependence. This interdependence is most dramatically evidenced by the effects of H_2-receptor antagonists and anticholinergic agents. H_2 blockers such as cimetidine inhibit the response not only to histamine but also to gastrin and cholinergic stimuli as well (11, 12). Anticholinergic agents inhibit the response to histamine and gastrin as well as to cholinergic stimuli (31, 32). Vagotomy and antrectomy are often used in combination in the surgical therapy of peptic ulcer disease. However, each of these surgical maneuvers when used alone inhibits the responses to all forms of stimulation, including stimulation by histamine and by food (32–35). Interactions can also be demonstrated by administering two of the three major secretagogues simultaneously; the responses to such combinations are frequently greater than the maximal response to the single agents (36). Interactions demonstrated in this manner are, however, species dependent and variable in magnitude (37).

Theories of Secretagogue Action and Interaction

The interdependence among gastric secretagogues has fascinated physiologists for several decades. These interactions are also of interest to the physician treating peptic ulcer disease, since the efficacy of available therapies results at least in part from effects on secretagogue interactions. Two major theories have been proposed to explain secretagogue interactions. Code (9) and Black (10) among others have championed a theory based on that initially proposed by MacIntosh (38), wherein histamine acts as the final effector for the action of all secretagogues on gastric secretion. According to this theory, gastrin and acetylcholine release histamine, and the parietal cell has functional receptors only for histamine. Studies to determine if histamine is released by gastrin and by acetylcholine have yielded conflicting results. The strongest evidence for an effect of gastrin on histamine metabolism is the induction of histidine decarboxylase by gastrin (16, 17), but this effect has been found only in the rat.

A second theory of the interactions between secretagogues is that the parietal cell has specific receptors for histamine, gastrin, and acetylcholine, and that potentiating interactions occur between these secretagogues at the parietal cell itself. It was tentatively proposed that these effects resulted from positive cooperative interactions, which occurred between binding sites for these agents (12); however, other possibilities now appear more likely. A third theory has been proposed: namely that the ability of H_2 blockers to inhibit gastrin and acetylcholine results from an antisecretory property of the drug not related to blockade of the H_2 receptor. This theory is unlikely in that the effect of H_2 blockers on histamine-stimulated acid secretion follows kinetics indicative of competitive antagonism. Also, these agents do not inhibit the action of acetylcholine or dibutyryl cyclic AMP on isolated parietal cells (see below). In the subsequent sections, studies of certain aspects of the cellular mechanisms involved in acid secretion are reviewed in an attempt to throw some light on the mode of action and interaction of gastric secretagogues.

CELLULAR EVENTS IN ACID SECRETION

This brief review of the cellular events in acid secretion concentrates on those steps that are likely to be rate-limiting regulatory steps in cellular response to stimulation. The discussion, following the format used at a recent conference on the parietal cell (39), is organized by cellular topography.

Events at the Basal Surface

HORMONE RECEPTOR INTERACTIONS In light of the two main theories of secretagogue interaction, the presence or absence of receptors for gastric secretagogues on the parietal cell is of considerable interest. This question can be studied by examining the binding of radioactively labeled hormones to parietal cells, or by determining the response of parietal cells to stimulation. Despite efforts to develop a system in which the binding of radioactively labeled gastrin to gastric mucosal cells can be measured, only one laboratory has reported positive results (40), and these remain controversial because of failure to adequately demonstrate biological specificity and sensitivity.

In an attempt to functionally demonstrate receptors, our laboratory sought techniques for the preparation of isolated parietal cells that would preserve the cells' responsiveness to stimulation. Canine mucosal cells were dispersed by sequential exposure to collagenase and EDTA, and the response of parietal cells to stimulation was determined by monitoring the consumption of oxygen (41), the uptake of ^{14}C-aminopyrine (42), and the morphological transformation (see below) (43). Indirect indices of parietal cell response must be used, since the secretion of acid by the apical cell membrane entails the simultaneous secretion of a bicarbonate ion at the basal cell membrane. The consumption of oxygen is tightly linked to the secretion of acid, a relation most clearly demonstrated in studies of the perfused, ex vivo canine stomach preparation (44). Aminopyrine, a compound used extensively as an index of mucosal blood flow (45), was first used as an index of parietal cell function in vitro by Berglindh et al (46) in their studies with isolated gastric glands. The uptake of ^{14}C-aminopyrine probably represents partition of the compound across a pH gradient. Aminopyrine is a weak base, with a pKa of 5.0, and thus, at physiological pH, the un-ionized form would exist in several-hundredfold excess over the ionized form. Aminopyrine in its un-ionized form at a pH of 7.4 is freely permeable. In contrast, at a pH near 2.0, aminopyrine would exist in a several-thousandfold excess in the ionized form, and therefore be very poorly permeable across lipophilic barriers. Thus with a pH gradient across a lipophilic barrier, aminopyrine would accumulate in the acid space.

Both aminopyrine uptake (42) and oxygen consumption (41, 47) show similar patterns for the actions and interactions of secretagogues on isolated parietal cells. Stimulation was seen with cholinergic agents, gastrin, histamine, and the phosphodiesterase inhibitor, isobutyl methyl xanthine (IMX). With this preparation of cells, gastrin and histamine were weak stimulants of the parietal cell. The response to histamine, however, was greatly enhanced by a concentration of IMX that itself caused no stimulation. These data indicate direct responses to gastrin, cholinergic agents, and histamine. It appears that the parietal cell has specific receptors for each of these agents in that the response to histamine was blocked by H_2 antagonists but not atropine, the response to carbamylcholine was blocked by atropine but not H_2 antagonists, and the small response to gastrin was blocked by neither inhibitor. Basal parietal cell activity was not affected by atropine or H_2 antagonists, which indicates that these isolated cells are not under resting-cholinergic or histaminic influences.

Since these data indicate that H_2 blockers and anticholinergic agents are specific in their actions on isolated parietal cells, they are at odds with the in vivo observations that show an apparent lack of specificity of these agents. This contradiction appears to be explained at least in part by secretagogue interactions (47). The response of isolated parietal cells to the combination of histamine and gastrin far exceeds the sum of the responses to the single agents. This enhancement is evident at histamine and gastrin concentrations as low as $10^{-7}M$ and $10^{-10}M$ respectively. Against this histamine-gastrin interaction, H_2 blockers show marked inhibition, leaving the residual response roughly equivalent to that produced by gastrin alone. Thus, if the gastrin response is enhanced by a background of histamine, then H_2 blockers will show an apparent inhibition of gastrin, an effect due to specific inhibition of histamine enhancement of gastrin action. A similar pattern of interaction was found between histamine and carbachol, with anticholinergics and H_2 antagonists retaining their specificity against the respective cholinergic and histamine components of the interactions. These agents showed an apparent lack of specificity in that, for example, when carbachol was studied against a histamine background, H_2 antagonists appeared to inhibit the carbachol response, an effect due to removal of histamine's enhancement of carbachol. From these data the apparent lack of specificity of the anticholinergics and H_2-receptor antagonists may be explained by effects of these agents on secretagogue interactions. In vivo, the parietal cell in the basal state may be exposed to threshold or subthreshold amounts of histamine, acetylcholine, and circulating gastrin. Therefore, the in vivo action of any single agent on the parietal cell may represent in fact an interaction between this agent and endogenous secretagogues. The data obtained with the isolated parietal cell by no means exclude the possibility that gastrin and acetylcholine may also have direct effects on the gastric histaminocyte stimulating histidine decarboxylase, histamine uptake, or histamine release.

Mediators of Hormone Action

Since the discovery of mediation of hormone action by cyclic AMP, there has been a great deal of interest in cyclic nucleotides as possible mediators for secretagogues that act on parietal cells. Although contradictory data abound in the literature regarding the possible role of cyclic nucleotides (48), there is an increasing body of data that suggests that histamine activates adenylate-cyclase activity in gastric fundic mucosa (49, 50). Studying dispersed mucosal cells (51), histamine (but not carbachol or gastrin) caused an increase in the production of cyclic AMP. With progressive enrichment of the parietal cells, through use of an elutriator rotor, it was possible to show a highly significant correlation between the parietal cell content and the histamine-induced rise in cyclic AMP production, which indicated that it was the parietal cell, and probably only the parietal cell, that accounted for this histamine effect.

Before one can seriously consider cyclic AMP as a candidate for mediating histamine action on the parietal cell, the effects of secretin and the prostaglandins on cyclic AMP production must be considered. Both secretin and the prostaglandins inhibit gastric acid secretion yet stimulate adenylate cyclase (52) and cause a rise in cyclic AMP production by unenriched mucosal cells (51). Using cell separation techniques, we found that the effects of secretin and prostaglandins on cyclic AMP production occurred primarily in cells other than the parietal cell (51). Furthermore, the effects of histamine on cyclic AMP production, oxygen consumption, and aminopyrine uptake occurred over a similar concentration range, which suggests that this effect of histamine on cyclic AMP may be correlated with the physiological response of the parietal cell to stimulation. The effects of prostaglandins and secretin on cyclic AMP were demonstrated at high concentrations (51, 52) and it remains unclear whether these effects are physiological or solely pharmacological.

There are no solid candidates for intracellular mediators of the effects of carbachol or gastrin. Studies with isolated parietal cells indicate that these effects are not mediated by cyclic AMP (51). Studies of the effects of cholinergic stimulation on cyclic GMP production yielded some positive results (49) in intact animals, although changes in cyclic GMP production were not found in dispersed mucosal cells stimulated with gastrin or carbachol (A. W. Wollin and A. H. Soll, unpublished observations).

Events in the Cytoplasm

METABOLIC EVENTS SUPPORTING GASTRIC SECRETION The sources of metabolic energy for gastric secretion are uncertain, although there is clearly an increase in oxygen consumption that is proportional to acid secretion and activation of the Krebs tricarboxylic acid cycle (53). There are conflicting studies on the source of the substrate pool. Studies in frog and rabbit gastric mucosa with labeled precursors suggest a preferential metabolism of fatty acids

(54), whereas in dog gastric mucosa major increases in glycolytic intermediates were found without significant changes in glycogen stores, which suggests metabolism of glucose (55). In the latter studies, increases in glycerol were also consistent with increased breakdown of triglycerides. The differences in these results may reflect enhancement of more than one metabolic pathway, with variations in the pattern under different experimental conditions and in different species.

A second major question is whether these metabolic events are a major primary effect of the gastric secretagogues in mobilizing substrates (thus constituting a rate-limiting metabolic step) or whether these events are secondary to the increase in demand for energy. With increased energy utilization or with uncouplers of oxidative phosphorylation such as dinitrophenol, the usual pattern is a shift of the respiratory-chain components to a more oxidized form (53). In contrast, in vitro studies with amphibian gastric mucosa indicate a transition of these respiratory-chain components to a more reduced form. These redox changes occur shortly after the onset of stimulation and before the actual secretion of acid commences. These observations combined with the increase in oxygen consumption are most consistent with a primary mobilization of substrate by secretagogues (53).

Two final related questions regard the source of the hydrogen ions for the secretion of acid and the means of delivery of metabolic energy to the hydrogen-ion pump. Two theories have been proposed to deal with these questions. The first theory, the redox theory, suggests that hydrogen ions are directly generated from the oxidation-reduction process, and delivered to the secretory surface via the electron transport system. The second theory, the ATP hypothesis, proposes that ATP generated from substrate metabolism provides the necessary energy via a specialized ATPase (see below).

Events at the Apical Surface

MORPHOLOGICAL TRANSFORMATION The parietal cell undergoes a very dramatic morphological transformation with stimulation (56). In the resting state, the cytoplasm of the parietal cell is filled with tubulovesicles, specialized smooth endoplasmic reticulum. With stimulation, these tubulovesicles coalesce to form extensive secretory canaliculi that are lined with microvilli and that communicate with the luminal surface of the cell. This transformation must involve the cytoskeleton of the cell, although the details of the role of microfilaments and microtubules in the transformation and in the equally complex reversion to the resting state remain unknown (56). The acid secretory apparatus is apparently housed in the membranes of the tubulovesicles and secretory canaliculi (see below), and therefore the dramatic morphological transformation that the parietal cell undergoes with stimulation serves to increase the luminal secretory surface.

THE H^+ PUMP Recent studies indicate that of the several ATPases present in gastric mucosa, a specialized K^+-stimulated ATPase represents a major component of the concentrating mechanism for the hydrogen ion (57). This K^+-stimulated ATPase has been found in abundance in five species thus far. It is associated with the parietal cells, and in developmental studies with tadpoles the appearance of this enzyme coincides with the development of the tubulovesicles and with appearance of the ability to secrete acid.

In recent years techniques have been developed for isolating, from gastric mucosa, vesicles derived apparently from the tubulovesicles of the parietal cell (55, 57). These vesicles contain the K^+-stimulated ATPase, and, when exogenous ATP is added, they are able to concentrate acid, as indicated by alkalinization of the medium.

Antibodies made to highly enriched preparations of these vesicles react with the supranuclear region of the parietal cell and inhibit both hydrogen-ion concentration by the vesicles and the K^+-ATPase (55). Studies with these vesicles indicate that behavior of the hydrogen-ion pump is best described by a model of active H^+/K^+ exchange, with a passive KCl limb (55). The function of these vesicles is not inhibited by Cl^- removal from the medium and therefore differs from in vitro studies with intact mucosa where removal of serosal Cl^- markedly inhibited acid secretion (58, 59). Under conditions of Cl^- removal in the latter studies with intact mucosa, hydrogen-ion secretion is an electrogenic mechanism, rather than a neutral exchange mechanism as it appears to be in isolated vesicles. The acid secretory mechanism thus appears to be more complex than simply the K^+-stimulated ATPase in the apical membrane. Nonetheless this enzyme appears to be a major component of the H^+ pump.

SUMMARY

The physiology of gastric acid secretion on the cellular level is complex and poorly understood; however, some advances have been made recently with the development of techniques for isolating parietal cells and membrane vesicles derived from parietal cells. From physiological studies with isolated cells, the parietal cell may have separate and specific receptors for histamine, gastrin, and acetylcholine in that (*a*) the response to histamine was inhibited by H_2 blockers but not by anticholinergics; (*b*) the response to carbachol was inhibited by anticholinergics but not H_2 blockers; and (*c*) the response to gastrin was inhibited by neither. Potentiating histamine-gastrin and histamine-carbachol interactions were observed; effects of anticholinergics and H_2-receptor antagonists on these interactions may account for the apparent lack of specificity displayed by these inhibitors in vivo. In studies with isolated cells and with intact mucosa, an increasing body of information indicates that histamine action may be mediated by generation of cyclic AMP. Mediators for gastrin and cholinergic stimulation remain unknown, as do the mechanisms for interactions between these agents. Studies with membrane vesicles derived from the parietal cell tubulovesicular apparatus indicate that a major component of the H^+-concentrating mechanism is a K^+-stimulated ATPase, although differences between the behavior of this isolated H^+ pump and that of intact mucosa do exist.

Since the H_2-histamine–receptor antagonists in vivo strongly inhibit all stimuli, there is little doubt of the major importance of histamine in the physiology of gastric acid secretion. However, little is known about which cells in gastric mucosa store histamine, or how it is formed and released.

Literature Cited

1. Reite, O. B. 1972. Comparative physiology of histamine. *Physiol. Rev.* 52:778–819
2. Grossman, M. I. 1978. Control of gastric secretion. In *Gastrointestinal Disease,* ed. M. H. Sleisenger, J. S. Fordtran. Philadelphia: Saunders. In press
3. Grossman, M. I. 1967. Neural and hormonal stimulation of gastric secretion of acid. In *Handbook of Physiology,* ed. C. F. Code, Sect. 6, Vol. 2, pp. 835–63. Washington, DC: Am. Physiol. Soc.
4. Roland, M., Berstad, A., Liavag, I. 1975. Effect of carbacholine and urecholine on pentagastrin-stimulated gastric secretion in healthy subjects. *Scand. J. Gastroenterol.* 10:357–63
5. Walsh, J. H., Grossman, M. I. 1975. Gastrin. *N. Engl. J. Med.* 292:1324–32, 1377–84
6. Ippoliti, A. F., Maxwell, V., Isenberg, J. I. 1976. Demonstration of the intestinal phase of gastric secretion in man. *Gastroenterology* 70: 896 (Abstr.)
7. Grossman, M. I., Brown, J. C., Adelson, J., Rothman, S .S., Lin, T. M., Chance, R. E., Gerring, E. L., Gregory, H., Creutzfeldt, W., Kokas, E., Nasset, E. S., Sasaki, H., Faloona, G. R., Unger, R. H., Andersson, S., Glass, G. B. J. 1974. Candidate hormones of the gut. *Gastroenterology* 67:730–55
8. Creutzfeldt, W. 1976. Effects of gastrointestinal hormones—physiological or pharmacological? In *Stimulus-Secretion Coupling in the Gastrointestinal Tract,* ed. R. M. Case, H. Goebell, pp. 415–28. Lancaster, England: MTP Press
9. Code, C. F. 1965. Histamine and gastric secretion; a later look, 1955–65. *Fed. Proc.* 24:1311–21
10. Black, J. W., Duncan, W. A. M., Durant, C. J., Ganellin, C. R., Parsons, E. M. 1972. Definition and antagonism of histamine H_2- receptors. *Nature* 236:385–90
11. Burland, W. L., Simkins, M. A. 1977. Cimetidine. *Proc. Int. Symp. Histamine H_2-Receptor Antagonists, 2nd.* Oxford: Excerpta Med. 392 pp.
12. Grossman, M. I., Konturek, S. J. 1974. Inhibition of acid secretion in dog by metiamide, a histamine antagonist acting on H_2 receptors. *Gastroenterology* 66:517–21
13. Lorenz, W., Schauer, A., Heitland, S., Calvoer, R., Werle, E. 1969. Biochemical and histochemical studies on the distribution of histamine in the digestive tract of man, dog and other mammals. *Naunyn-Schmiedebergs Arch. Pharmakol.* 265:81–100
14. Cross, S. A. M., Ewen, S. W. B., Rost, F. W. D. 1971. A study of the methods available for the cytochemical localization of histamine by fluorescence induced with *o*-phthalaldehyde or acetaldehyde. *Histochem. J.* 3:471–76
15. Beaven, M. A. 1976. Histamine. *N. Engl. J. Med.* 294:30–36, 320–25
16. Beaven, M. A. 1978. Histamine: a reassessment of its role in physiological and pathological processes. *Prog. Allergy.* In press
17. Kahlson, G., Rosengren, E. 1968. New approaches to the physiology of histamine. *Physiol. Rev.* 48:155–96
18. Debas, H. T., Grossman, M. I. 1975. Chemicals bathing oxyntic gland area stimulate acid secretion in dog. *Gastroenterology* 69:654–59
19. Isenberg, J. I., Maxwell, V. 1977. Intravenous infusion of amino acids stimulates gastric acid secretion in man. *Gastroenterology* 72:1074 (Abstr.)
20. Konturek, S. J. 1976. Somatostatin and gastrointestinal secretions. *Scand. J. Gastroenterol.* 11:1–4
21. Pearse, A. G. E., Polak, J. M., Bloom, S. R. 1977. The newer gut hormones: cellular sources, physiology, pathology, and clinical aspects. *Gastroenterology* 72:746–61
22. Robert, A. 1976. Antisecretory, antiulcer, cytoprotective and diarrheogenic properties of prostaglandins. In *Advances in Prostaglandin and Thromboxane Research,* ed. B.

Samuelsson, R. Paoletti, Vol. 2, pp. 507–20. New York: Raven

23. Main, I. H. M., Whittle, B. J. R. 1976. The role of prostaglandins in gastric acid secretion. See Ref. 8, pp. 147–57
24. Konturek, S. J., Biernat, J., Kwiecień, N., Oleksy, J. 1975. Effect of gluacagon on meal-induced gastric secretion in man. *Gastroenterology* 68:448–54
25. Valenzuela, J. E. 1976. Dopamine as a possible neurotransmitter in gastric relaxation. *Gastroenterology* 71:1019–22
26. Jaffe, B. M., Kopen, D. F., Lazan, D. W. 1977. Endogenous serotonin in the control of gastric acid secretion. *Surgery* 82:156–63
27. Pederson, R. A., Brown, J. C. 1972. Inhibition of histamine-, and pentagastrin-, and insulin-stimulated canine gastric secretion by pure "gastric inhibitory polypeptide." *Gastroenterology* 62:393–400
28. Bugat, R., Thompson, M. R., Aures, D., Grossman, M. I. 1976. Gastric mucosal lesions produced by intravenous infusion of aspirin in cats. *Gastroenterology* 71:754–59
29. Vane, J. R. 1971. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nature New Biol.* 231:232–35
30. Hotz, J., Goebell, H., Ziegler, R. 1976. The role of calcium in gastric acid and pepsin secretion. See Ref. 8, pp. 159–72
31. Code, C. F., Hightower, N. C., Hallenbeck, G. A. 1951. Comparison of the effects of methantheline bromide (Banthine) and atropine on the secretory responses of vagally innervated and vagally denervated gastric pouches. *Gastroenterology* 19:254–64
32. Konturek, S. J., Wysocki, A., Oleksy, J. 1968. Effect of medical and surgical vagotomy on gastric response to graded doses of pentagastrin and histamine. *Gastroenterology* 54:392–400
33. Greenall, M. J., Lyndon, P. J., Goligher, J. C., Johnston, D. 1975. Long term effect of highly selective vagotomy on basal and maximal acid output in man. *Gastroenterology* 68:1421–25
34. Olbe, L. 1964. Effect of resection of gastrin-releasing regions on acid response to sham feeding and insulin hypoglycemia in Pavlov pouch dogs. *Acta Physiol. Scand.* 62:169–75
35. Bergegardh, S., Broman, G., Knutson, U., Palmer, L., Olbe, L. 1976. Gastric acid responses to graded i.v. infusions of pentagastrin and histalog in peptic ulcer patients before and after antrum-bulb resection. *Scand. J. Gastroenterol.* 11:337–46
36. Johnson, L. R., Grossman, M. I. 1969. Potentiation of gastric acid response in the dog. *Gastroenterology* 56:687–92
37. Brooks, A. M., Johnson, L. R., Grossman, M. I. 1970. Effects of combinations of histamine and pentagastrin on gastric secretion in man and dog. *Gastroenterology* 58:470–75
38. MacIntosh, F. C. 1938. Histamine as a normal stimulant of gastric secretion. *Q. J. Exp. Physiol.* 28:87–98
39. Grossman, M. I., ed. 1977. Conference on the oxyntic cell. *Gastroenterology* 73:873–970
40. Soumarmon, A., Cheret, A. M., Lewin, M. J. M. 1977. Localization of gastrin receptors in intact isolated and separated fundic rat cells. *Gastroenterology* 73:900–3
41. Soll, A. H. 1978. The actions of secretagogues on oxygen uptake by isolated mammalian parietal cells. *J. Clin. Invest.* In press
42. Soll, A. H. 1977. Secretagogue stimulation of O$_2$ consumption and ^{14}C-aminopyrine uptake by enriched canine parietal cells. *Gastroenterology* 72:1166 (Abstr.)
43. Soll, A. H., Lechago, J., Walsh, J. H. 1976. The isolated mammalian parietal cell: morphological transformation induced by secretagogues. *Gastroenterology* 70:975 (Abstr.)
44. Kowalewski, K., Kolodej, A. 1972. Relation between hydrogen ion secretion and oxygen consumption by *ex vivo* isolated canine stomach, perfused with homologous blood. *Can. J. Physiol. Pharmacol.* 50:955–61
45. Jacobson, E. D., Linford, R. H., Grossman, M. I. 1966. Gastric secretion in relation to mucosal blood flow studied by a clearance technic. *J. Clin. Invest.* 45:1–13
46. Berglindh, T., Helander, H. F., Obrink, K. J. 1976. Effects of secretagogues on oxygen consumption, aminopyrine accumulation, and

morphology in isolated gastric glands. *Acta Physiol. Scand.* 97: 401–14

47. Soll, A. H. 1978. The interaction of histamine with gastrin and carbamylcholine on oxygen uptake by isolated mammalian parietal cells. *J. Clin. Invest.* In press
48. Jacobson, E. D., Thompson, W. J. 1976. Cyclic AMP and gastric secretion: the illusive second messenger. *Adv. Cyclic Nucleotide Res.* 7: 199–224
49. Bieck, P. R. 1976. The role of cyclic nucleotides in gastric acid secretion. See Ref. 8, pp. 129–45
50. Dousa, T. P., Dozois, R. R. 1977. Interrelations between histamine, prostaglandins and cyclic AMP in gastric secretion: a hypothesis. *Gastroenterology* 73:904–12
51. Soll, A. H., Wollin, A. 1977. The effects of histamine, prostaglandin E_2, and secretin on cyclic AMP in separated canine fundic mucosal cells. *Gastroenterology* 72:1166 (Abstr.)
52. Thompson, W. J., Chang, L. K., Rosenfeld, G. C., Jacobson, E. D. 1977. Activation of rat gastric mucosal adenylyl cyclase by secretory inhibitors. *Gastroenterology* 72:251–54
53. Hersey, S. J. 1974. Interactions between oxidative metabolism and acid secretion in gastric mucosa. *Biochim. Biophys. Acta* 344: 157–203
54. Hersey, S. J. 1977. Metabolic changes associated with gastric stimulation. *Gastroenterology* 73: 914–19
55. Sachs, G., Chang, H., Rabon, E., Shackman, R., Sarau, H. M., Saccomani, G. 1977. Metabolic and membrane aspects of gastric H^+ transport. *Gastroenterology* 73:931–40
56. Forte, T. M., Machen, T. E., Forte, J. G. 1977. Untrastructural changes in oxyntic cells associated secretory function: a membrane-recycling hypothesis. *Gastroenterology* 73:941–55
57. Forte, J. G., Lee, H. C. 1977. Gastric ATPases: a review of their possible role in HCl secretion. *Gastroenterology* 73:921–26
58. Durbin, R. P. 1977. Chloride transport and acid secretion in stomach. *Gastroenterology* 73:927–30
59. Rehm, W. S., Sanders, S. S. 1977. Electrical events during activation and inhibition of gastric HCl secretion. *Gastroenterology* 73:959–69.

Ann. Rev. Med. 1978. 29:509–20
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CONTROL OF THE ONSET OF PUBERTY

◆7295

Robert M. Boyar, M.D.
Departments of Internal Medicine and Pediatrics, University of Texas Health Science Center, Dallas, Texas 75235

INTRODUCTION

Recent advances in our understanding of the endocrine concomitants of sleep and the episodic character of pituitary hormone secretion have led to new insights concerning the importance of the central nervous system (CNS) in the control of the onset of human puberty (1). Human puberty is a complex process that produces significant changes in the psychological, physical, and reproductive state of the individual. The processes that eventuate in the onset and progression of human puberty are not completely understood; however, significant progress has been made in our understanding of some of the hormonal concomitants of this process. It was recently established that at some point in "late" prepuberty, an augmentation of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) occurs synchronously with sleep (2–4). This sleep-related gonadotropin program initiates testosterone secretion in boys (5) and estradiol secretion in girls (6), which ultimately result in the clinical characteristics associated with human puberty. The mechanisms responsible for the initiation of this pubertal program have not been identified; however, this review describes important advances in our understanding of possible factors.

NEUROENDOCRINE CONTROL OF PUBERTY

The measurement of plasma LH and FSH at 20-min intervals for 24 hr in prepubertal, pubertal, and adult subjects showed an ontogenetic change both in the pulsatile nature of the hormone secretory patterns and in their relationship to the sleep-wake cycle. Early prepubertal children have low LH and FSH levels throughout the 24-hr sleep-wake cycle with only minimal evidence of pulsatile secretory activity. At some time in late prepuberty, a marked augmentation of LH secretion (and to a lesser extent FSH secretion) occurs synchronously with polygraphically monitored sleep (Figure 1). The relationship of this nocturnal gonadotropin program to sleep was shown clearly by delayed sleep onset studies

0066-4219/78/0401-0509$01.00

and sleep reversal studies. These studies showed that the augmented LH secretion shifted to the time of the new sleep period (7). This augmentation of LH secretion could not be reproduced by blindfolding pubertal subjects while awake.

As puberty progresses, pulsatile LH secretion begins to occur during waking, but the levels of LH are still higher during the sleep period. At the completion of sexual maturation, there is no longer a consistent difference between LH secretory activity during sleep and waking periods (8). Although random LH and FSH values in adults can overlap those of prepubertal subjects, the 24-hr mean LH concentration in adults is approximately twice that of prepubertal children. Of interest is the fact that the testosterone concentration increases approximately 20-fold compared to the two fold increase in the mean LH concentration. These results suggest that a change in the sensitivity of the hypothalamic-pituitary gonadotropin axis is an additional concomitant of the normal pubertal process.

Since the recurrent nonREM-REM sleep cycle is probably generated by sequential changes in neurotransmitter function (9), and the control of luteinizing hormone–releasing hormone (LH-RH) is almost certainly under the control of monoaminergic imputs, it should not have been totally unexpected to find a relationship between sleep and gonadotropin secretion. In primates, α-adrenergic mechanisms appear to be important in the origin of pulsatile LH secretion in castrate monkeys, as phentolamine, an α-adrenergic blocker, completely suppresses pulsatile LH secretory activity (10). Recently, Leblanc et al

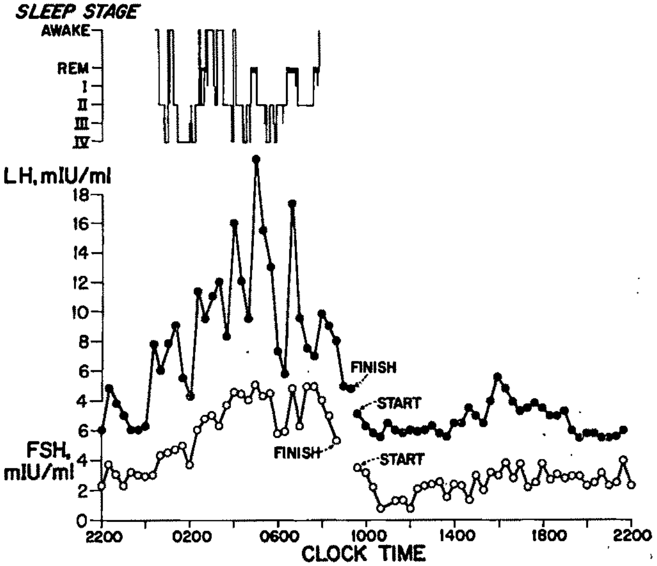


Figure 1 Plasma LH and FSH concentrations sampled at 20-min intervals for 24 hr in a normal early pubertal girl.

(11) showed that dopamine infusion suppressed LH secretion in normal subjects, which suggests an inhibitory role for this neurotransmitter in the control of LH secretion. Since dopamine receptors are present on the pituitary, and dopamine does not cross the blood-brain barrier, the effect may not be the result of an alteration of hypothalamic dopamine levels. Further studies will be needed to resolve this point.

THE ROLE OF ADRENAL ANDROGENS DURING PUBERTY

Adrenal androgen concentrations (i.e. DHEA, DHEA-S, androstenedione) are low during early prepuberty. A role for these steroids in the maintenance of the exquisite sensitivity of the hypothalamic-pituitary axis during prepuberty has been suggested. Recently, De Peretti & Forest (12) showed that basal and ACTH-stimulated DHEA levels are high between 2 and 5 months of age and between 6 and 14 years of age, prior to reaching adult levels. During early prepuberty, basal DHEA levels are low and fail to increase after ACTH stimulation. These changes in DHEA parallel the changes in the sensitivity of the hypothalamic-pituitary axis to the inhibitory effects of gonadal steroids. Of additional interest is the finding that adrenal DHEA responsiveness to ACTH returns at approximately the same time as the initiation of the pubertal LH secretory program. This parallelism suggests a possible effect of adrenal androgens, particularly DHEA, on the initiation of this pubertal program or an effect on the inhibitory CNS mechanisms that restrain the initiation of this program. This study also showed that the rise in DHEA was due to heightened adrenal responsiveness, as the DHEA concentrations after ACTH stimulation were always greater than those after HCG stimulation. The factor(s) responsible for this heightened adrenal responsiveness to ACTH is unknown; however, recent studies suggest that prolactin may be important (13).

In 1973, we reported that two boys with untreated congenital adrenal hyperplasia (CAH) showed LH secretory programs that were identical to normal older pubertal boys. This finding suggested that the prolonged exposure to elevated concentrations of adrenal androgens resulted in the premature initiation of the normal pubertal LH program (3). This observation was also consistent with the clinical observation that children with CAH who remain untreated for prolonged periods rapidly enter clinical puberty after commencing cortisol replacement therapy.

The mechanism by which adrenal androgens result in the initiation of CNS puberty is unknown. Since androstenedione can be peripherally converted in fat (14) as well as in brain (15) to estrone, it is possible that the increased levels of biologically active estrogens may be the critical factor. This is supported by the "regression" or "arrest" of the LH secretory program that is found in anorexia nervosa patients who show a decrease in plasma estrogens (16). Further studies of extraglandular aromatization during puberty should be performed to determine if the rate of aromatization also changes during puberty. It is possible that increased substrate (adrenal androgens) and increased rates of aromatization are both responsible for the alteration in hypothalamic sensitivity that results in the initiation of the pubertal LH program.

THE ROLE OF LUTEINIZING HORMONE-RELEASING HORMONE

The onset of pulsatile LH secretion in pubertal subjects during sleep is most probably the result of the pulsatile secretion of LH-RH. Measurements of LH-RH in portal blood of primates (17) and rodents (18) showed that this hypothalamic peptide is also secreted episodically. Measurements of LH-RH activity in peripheral blood of humans also showed a pulsatile secretory pattern (19). Recently, studies of radioimmunoassayable LH-RH showed that levels of this peptide in peripheral blood are extremely low (20). This finding coupled with the knowledge that this peptide has an extremely rapid half-life in plasma suggests that it may be difficult to gain useful information in humans by utilizing radioimmunoassay techniques for plasma LH-RH.

The synthesis of LH-RH made it possible for investigators to study its effect on the release of LH and FSH during sexual maturation. Studies showed that prepubertal children release only minimal amounts of LH; however, there is a progressive increase in LH responsiveness as puberty progresses (21–23). This probably reflects the gradual increase in pituitary LH stores as a result of an increase in endogenous LH-RH during sleep. Of additional interest is the observation that prepubertal girls release more FSH than prepubertal or

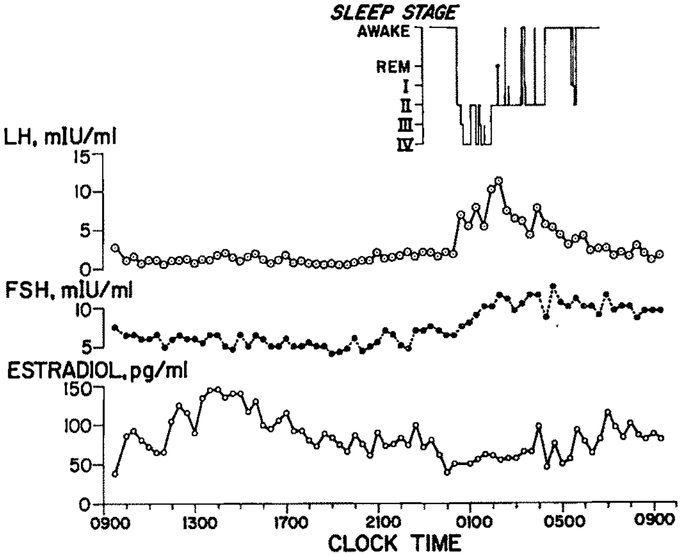


Figure 2 Plasma LH, FSH, and estradiol 24 hr secretory patterns in a normal pubertal girl.

pubertal boys at any stage of maturation. The mechanism responsible for this sex difference in FSH responsiveness to LH-RH is not known. It does not appear to be the result of a difference in gonadal sex steroid levels since these are equivalent in prepubertal children of both sexes. It is possible that the levels of "inhibin," the "putative" nonsteroidal FSH inhibitor, may be lower in prepubertal girls, and this would account for the sex difference in FSH responsiveness to LH-RH.

The availability of LH-RH for clinical use has provided new information concerning the etiology of idiopathic hypogonadotropic hypogonadism (IHH). In some patients with this disorder, the LH and FSH response to a single infusion of LH-RH may be impaired; however, repetitive administration of LH-RH will result in a normalization of the response (24). The same phenomenon is also seen in anorexia nervosa patients, who show impaired responses to single infusions of LH-RH, but normalize their response after repetitive LH-RH administration (25). Nillius & Wide (26) reported the induction of ovulation in patients with anorexia nervosa with chronic LH-RH administration. Finally, the observation that chronic self-administration of LH-RH can result in the initiation of puberty and normal spermatogenesis in patients with IHH is further evidence supporting the role of LH-RH in the initiation of normal puberty (27).

GONADOTROPIN-ESTRADIOL SECRETORY PATTERNS IN PUBERTAL GIRLS

Studies of the 24-hr LH and FSH secretory patterns in pubertal girls showed that simultaneous augmentation of LH and FSH during sleep occurred more consistently in pubertal girls than in pubertal boys. The reason for this was not readily apparent until studies of the 24-hr LH, FSH, and estradiol secretory patterns in pubertal girls were completed (Figure 2). The results of these studies showed that plasma estradiol levels were inhibited during the period of maximal gonadotropin secretion and then increased 10–12 hr later during the day, a period of minimal gonadotropin secretion (6). This lowering of plasma estradiol during sleep permits FSH secretion to occur synchronously with LH secretion. In pubertal boys, testosterone secretion is augmented simultaneously with LH secretion during sleep, which results in suppression of FSH secretion in mid-late pubertal boys (5).

The mechanism(s) responsible for the delayed rise in estradiol levels and the suppression of estradiol secretion during sleep was clarified by recent studies conducted in rats. Katz & Armstrong (28) showed that LH inhibits estradiol secretion in vivo by decreasing the activity of the aromatizing enzyme system. Also, Dorrington et al (29) performed in vitro studies that showed that granulosa cells from hypophysectomized rats can aromatize testosterone in the presence of highly purified FSH, but not in the presence of LH alone. These investigators proposed that FSH stimulates the aromatization of testosterone in isolated granulosa cells. They concluded that a two-cell system is necessary for the synthesis of estradiol. Firstly, the theca interna cells synthesize precursor androgens under the influence of LH, and secondly, these androgens are aromatized to estradiol in granulosa cells under the influence of FSH. These results provide a possible explanation of the time lag in estradiol secretion following gonadotropin stimulation reported to occur in pubertal girls.

PUBERTAL MECHANISMS IN GONADAL DYSGENESIS

Studies of LH and FSH levels in children with gonadal dysgenesis provide some insights concerning the inhibitory mechanisms that restrain gonadotropin secretion during prepuberty. Conte et al (30) showed that children with gonadal dysgenesis have elevated FSH levels (compared with age-matched controls) between birth and three years of age and again between 10 and 16 years of age. Between 6 and 10 years of age, the FSH levels are restrained so they are only minimally above age-matched controls. The mechanism responsible for the reappearance of gonadotropin restraint is not known; however, it is probably mediated by CNS inhibitory factors since these patients have no functioning gonadal tissue. Odell & Swerdloff (31) reported that agonadal subjects showed CNS-pituitary-axis maturation at approximately the same age as their peers with normal gonadal function. These data suggest that gonadal factors may not be the critical factor in the initiation of the increased gonadotropin secretion at the onset of CNS puberty.

In studies of the 24-hr LH and FSH secretory patterns in gonadal dysgenesis, there is a preservation of the sleep-associated increase in LH and FSH secretion (Figure 3). Pulsatile secretion is evident during waking. However, with the onset of sleep, there is a marked augmentation of LH and FSH secretion, similar qualitatively to the secretory patterns of normal pubertal girls (32). Twenty-four hour LH and FSH secretory patterns in adult patients with gonadal dysgenesis

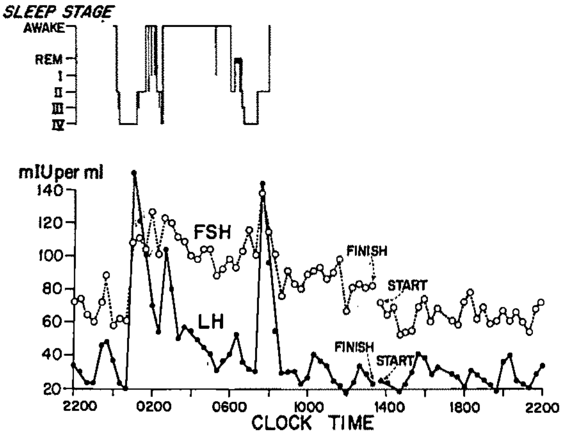


Figure 3 Plasma LH and FSH concentrations sampled at 20-min intervals for 24 hr in the 12-yr-old girl with gonadal dysgenesis. Above the period of nocturnal sleep, a histogram shows the sequence of sleep stages.

show random LH and FSH secretory activity throughout the 24-hr period, similar to normal adult women (32, 33). These observations suggest that the onset and progression of the pubertal LH program in patients without functioning ovarian tissue is preserved qualitatively, though the levels of the gonadotropins are much higher. Therefore, gonadal factors play a role in the regulation of the quantitative aspects of LH and FSH secretion but not the qualitative characteristics of this nycterohemeral rhythm.

THE METABOLISM OF TESTOSTERONE DURING PUBERTY

Recently, Zumoff et al (34) showed that the metabolism of exogenously administered testosterone varied during sexual maturation. The results of these studies showed that prepubertal children of both sexes had similar androsterone (A): etiocholanolone (E) ratios. In pubertal subjects, the A:E ratio decreased significantly, especially in the pubertal girls. This sex difference was highly significant. Treatment of two sexually immature young adult girls with estrogen and progesterone resulted in a significant decrease of their A:E ratio, similar to that observed in normal pubertal girls. These results suggest that the estrogen-progesterone hormonal milieu of pubescent girls is associated with a change in the metabolism of exogenously administered testosterone. This change is consistent with a decrease in the amount of testosterone that is available for 5α-reduction by the liver.

STUDIES OF NEGATIVE AND POSITIVE FEEDBACK DURING PUBERTY

Studies by Kelch and co-workers (35) showed that prepubertal children exhibit uniform suppresion of their urinary FSH levels after 5μg of ethinyl estradiol, but pubertal children who received 10μg ethinyl estradiol exhibited no suppression of urinary FSH levels. These findings were felt to be consistent with a decrease in the sensitivity of the hypothalamic-pituitary gonadotropin axis to sex steroid suppression as children went from prepuberty to puberty. These findings would also be consistent with a change in the efficiency of ethinyl estradiol metabolism. Pubertal children may metabolize this steroid more efficiently than prepubertal children, which would result in less steroid available for hypothalamic-pituitary suppression. This point could be resolved with measurements of plasma estradiol in children of differing pubertal stages receiving ethinyl estradiol.

Additional evidence supporting a change in hypothalamic-pituitary mechanisms during sexual maturation is the finding that positive feedback is not acquired in girls until mid-to-late puberty. Reiter et al (36) showed that prepubertal girls fail to exhibit an LH increase following an initial fall after 17β-estradiol administration. These results contrasted with those of three older girls (who were either of pubertal age or had clinical puberty) who showed LH rises after 17β-estradiol administration. These investigators concluded that the capacity to respond to 17β estradiol with a rise in LH is acquired sometime in mid-to-late puberty.

DISORDERS OF PUBERTY

Anorexia Nervosa

Recent studies of LH secretory dynamics in patients with anorexia nervosa showed a regression or arrest of the LH secretory program to one that is similar to prepubertal or pubertal children (16). These findings suggested that some factor associated with the loss of body weight concomitantly affected the release of LH. Improved caloric intake and nutritional status resulted in normalization of the LH secretory program, which further supported this view (Figure 4). Since repetitive LH-RH stimulation in patients with anorexia nervosa can result in normalization of pituitary responsiveness to LH-RH and ovulation without any change in body weight, the pituitary and ovarian abnormality must be secondary to chronic LH-RH deficiency. The primary abnormality that results in decreased LH-RH synthesis and release has not been identified. Whether it is a hypothalamic abnormality secondary to the neurochemical concomitants of the psychiatric disorder and/or the severe malnutrition remains to be elucidated. In any event, this disorder provides a model for the study of factors that may be relevant to the regulation of gonadotropin secretion during puberty.

Hypogonadotropic Hypogonadism

Idiopathic hypogonadotropic hypogonadism (IHH) is a disorder primarily occurring in males who show a persistent failure to initiate normal puberty (Figure 5). These patients often have anosmia, hyposmia, cleft palate, microphallus, and cryptorchidism. The microphallus and cryptorchidism are probably the result of the inadequate release of LH during the second and the third trimester of gestation, a time when testicular descent and phallic growth occur. The use of LH-RH for both acute and chronic administration has shown that this is a heterogeneous disorder almost certainly of hypothalamic origin (37). The availability of LH-RH for chronic administration should enable these patients to initiate and complete normal puberty as well as initiate and maintain normal spermatogenesis.

CONCLUSION

This review has made an attempt to delineate some of the factors that may play a role in the control of the onset of human puberty. It is evident that puberty is not a climactic event, but more likely a slowly evolving process initiated early in life. The finding of augmented gonadotropin secretion occurring synchronously with sleep provides a new biologic index for identifying CNS puberty. This marker of CNS puberty should enable investigators to interpret changes in adrenal androgen secretion, LH-RH responsiveness, testosterone metabolism, body composition, and positive and negative feedback in relation to this specific CNS

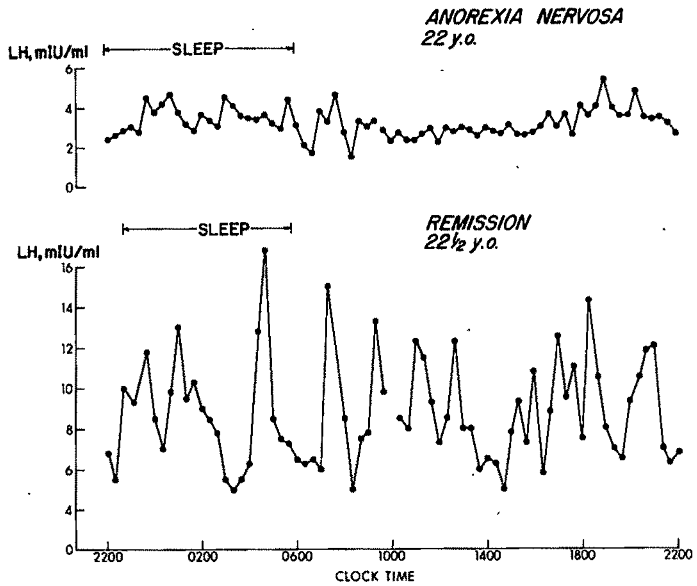


Figure 4 Plasma LH and FSH concentrations every 20 min for 24 hr in a patient with anorexia nervosa and secondary amenorrhea before and after clinical improvement.

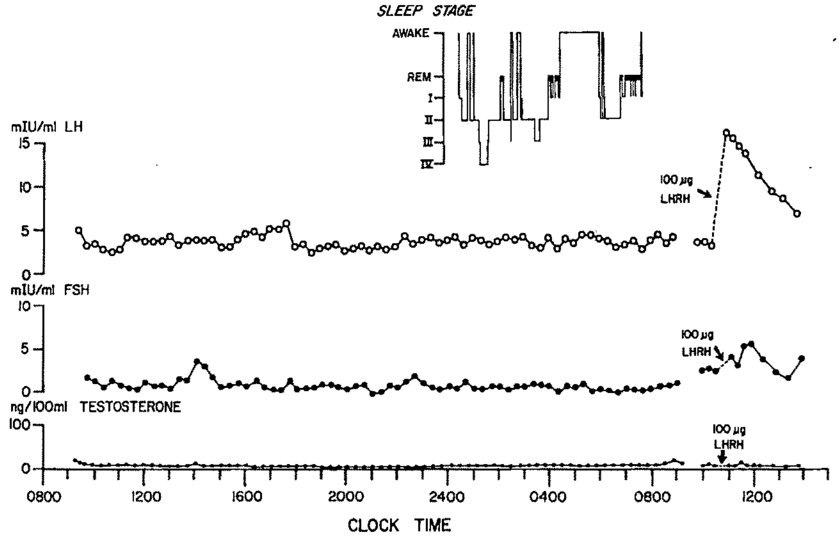


Figure 5 The 24-hr LH, FSH, and testosterone levels before and after 100μg LH-RH in a patient with isolated gonadotropin deficiency. The sleep histogram is depicted above the normal nocturnal sleep period.

event. The future availability of synthetic LH-RH for chronic administration to patients with isolated gonadotropin deficiency will enable investigators to delineate precisely the sequential changes in the gonadotropin-gonadal axis in response to chronic LH-RH administration. It will also provide physiologic replacement therapy for patients with this disorder. Future studies of the endocrine concomitants of human puberty should continue to provide new insights into this important biological process.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the KROC Foundation and from the Clinical Centers Branch (5-MO1-RR-00633-05). The author wishes to thank Kay Harper for secretarial assistance.

Literature Cited

1. Weitzman, E. D., Boyar, R. M., Kapen, S., Hellman, L. 1975. The relationship of sleep stages to neuroendocrinic secretion and biological rhythms in man. *Recent Prog. Horm. Res.* 31:399–441
2. Boyar, R., Finkelstein, J., Roffwarg, H., Kapen, S., Weitzman, E., Hellman, L. 1972. Synchronization of augmented luteinizing hormone secretion with sleep during puberty. *N. Engl. J. Med.* 287:582–86
3. Boyar, R. M., Finkelstein, J. W., David, R., Roffwarg, H., Kapen, S., Weitzman, E. D., Hellman, L. 1973. Twenty-four hour patterns of plasma luteinizing hormone and follicle-stimulating hormone in sexual precocity. *N. Engl. J. Med.* 289:282–86
4. Judd, H. M., Parker, D. C., Siler, T. M., Yen, S. S. C. 1974. The nocturnal rise of plasma testosterone in pubertal boys. *J. Clin. Endocrinol. Metab.* 38:710–13
5. Boyar, R. M., Rosenfeld, R. S., Kapen, S., Finkelstein, J. W., Roffwarg, H. P., Weitzman, E. D., Hellman, L. 1974. Human puberty: simultaneous augmented secretion of luteinizing hormone and testosterone during sleep. *J. Clin. Invest.* 54:609–18
6. Boyar, R. M., Wu, R. H. K., Roffwarg, H., Kapen, S., Weitzman, E. D., Hellman, L., Finkelstein, J. W. 1976. Human puberty: twenty-four-hour estradiol patterns in pubertal girls. *J. Clin. Endocrinol. Metab.* 43:1418–22
7. Kapen, S., Boyar, R. M., Finkelstein, J. W., Hellman, L., Weitzman, E. D. 1974. Effect of sleep-wake cycle reversal on luteinizing hormone secretory pattern in puberty. *J. Clin. Endocrinol. Metab.* 39:293–99
8. Boyar, R. M., Perlow, M., Hellman, L., Kapen, S., Weitzman, E. 1972. Twenty-four hour pattern of luteinizing hormone secretion in normal men with sleep stage recording. 1972. *J. Clin. Endocrinol. Metab.* 35:73–81
9. Jouvet, M. 1969. Biogenic amines and the states of sleep. *Science* 163:32–41
10. Knobil, E. 1975. On the control of gonadotropin secretion in the Rhesus monkey. *Recent Prog. Horm. Res.* 30:1–36
11. Leblanc, H., Lachelin, G. C. L., Abu-Fadil, S., Yen, S. S. C. 1976. Effects of dopamine infusion on pituitary hormone secretion in humans. *J. Clin. Endocrinol. Metab.* 43:668–74
12. De Peretti, E., Forest, M. G. 1976. Unconjugated dehydroepiandrosterone plasma levels in normal subjects from birth to adolescence in human: The use of a sensitive radioimmunoassay. *J. Clin. Endocrinol. Metab.* 43:982–91
13. Bassi, F., Giusti, G., Borsi, L., Cattaneo, S., Giannotti, P., Forti, G., Pazzagli, M., Vigiani, C., Serio, M. 1977. Plasma androgens in women with hyperprolactinaemic amenorrhea. *Clin. Endocrinol.* 6:5–10
14. Grodin, J. M., Siiteri, P. K., Mac-

Donald, P. C. 1973. Source of estrogen production in the postmenopausal women. *J. Clin. Endocrinol. Metab.* 36:207–14
15. Naftolin, F., Ryan, K. J., Petro, Z. 1971. Aromatization of androstenedione by the diencephalon. *J. Clin. Endocrinol. Metab.* 33:368–70
16. Boyar, R. M., Katz, J., Finkelstein, J. W., Kapen, S., Weiner, H., Weitzman, E. D., Hellman, L. 1974. Anorexia nervosa: Immaturity of the 24-hour luteinizing hormone secretory pattern. *N. Engl. J. Med.* 291:861–65
17. Carmel, P. W., Araki, S., Ferin, M. 1976. Pituitary stalk portal blood collection in rhesus monkeys: evidence for pulsatile release of gonadotropin-releasing hormone (GRH). *Endocrinology* 99:243–48
18. Sarkar, D. K., Chiappa, S. A., Fink, G. 1976. Gonadotropin-releasing hormone surge in prooestrous rats. *Nature* 264:461–63
19. Seyler, L. E., Reichlin, S. 1973. Luteinizing hormone-releasing factor (LRF) in plasma of postmenopausal women. *J. Clin. Endocrinol. Metab.* 37:197–203
20. Mortimer, C. G., McNeilly, A. S., Rees, L. H., Lowry, P. J., Gilmore, D., Dobbie, H. G. 1976. Radioimmunoassay and chromatographic similarity of circulating endogenous gonadotropin-releasing hormone and hypothalamic extracts in man. *J. Clin. Endocrinol. Metab.* 43:882–88
21. Roth, J. C., Kelch, R. P., Kaplan, S. L., Grumbach, M. M. 1972. FSH and LH response to luteinizing hormone–releasing factor in prepubertal and pubertal children, adult males and patients with hypogonadotropic and hypergonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* 35:926–30
22. Job, J. C., Garnier, P. E., Chaussain, J. L., Milhaud, G. 1972. Elevation of serum gonadotropins (LH and FSH) after releasing hormone (LH-RH) injection in normal children and in patients with disorders of puberty. *J. Clin. Endocrinol. Metab.* 35:473–76
23. Grumbach, M. M., Roth, J. C., Kaplan, S. L., Kelch, R. P. 1974. In *Control of the Onset of Puberty,* ed. M. M. Grumbach, G. D. Grave E. F. Mayer, pp. 115–66. New York: Wiley. 484 pp.
24. Reitano, J. F., Caminos-Torres, R., Snyder, P. J. 1975. Serum LH and FSH responses to the repetitive administration of gonadotropin-releasing hormone in patients with idiopathic hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* 41:1035–42
25. Yoshimoto, Y., Moridera, K., Imura, H. 1975. Restoration of normal pituitary gonadotropin reserve by administration of luteinizing-hormone-releasing hormone in patients with hypogonadotropic hypogonadism. *N. Engl. J. Med.* 292:242–45
26. Nillius, S. J., Wide, L. 1975. Gonadotrophin-releasing hormone treatment for induction of follicular maturation and ovulation in amenorrhoeic women with anorexia nervosa. *Br. Med. J.* 3:405–8
27. Mortimer, C. H., McNeilly, A. S., Fisher, R. A., Murray, M. A. F., Besser, G. M. 1974. Gonadotrophin-releasing hormone therapy in hypogonadal males with hypothalamic or pituitary dysfunction. *Br. Med. J.* 4:617–21
28. Katz, Y., Armstrong, D. T. 1976. Inhibition of ovarian estradiol-17β secretion by luteinizing hormone in prepubertal, pregnant mare serum-treated rats. *Endocrinology* 99:1442–47
29. Dorrington, J. H., Moon, Y. S., Armstrong, D. T. 1975. Estradiol-17β biosynthesis in cultured granulosa cells from hypophysectomized immature rats; stimulation by follicle-stimulating hormone. *Endocrinology* 97:1328–30
30. Conte, F. A., Grumbach, M. M., Kaplan, S. L. 1975. A diaphasic pattern of gonadotropin secretion in patients with the syndrome of gonadal dysgenesis. *J. Clin. Endocrinol. Metab.* 49:670–74
31. Odell, W. D., Swerdloff, R. S. 1976. Etiologies of sexual maturation: A model system based on the sexually mature rat. *Recent Prog. Horm. Res.* 32:245–77
32. Boyar, R. M., Finkelstein, J. W., Roffwarg, H., Kapen, S., Weitzman, E. D., Hellman, L. 1973. Twenty-four hour luteinizing hormone secretory patterns in gonadal dysgenesis. *J. Clin. Endocrinol. Metab.* 37:521–25
33. Boyar, R. M., Wu, R. H. K., Liao,

P. Y., Finkelstein, J. W. 1977. Endocrine and cytogenetic studies in a patient with Turner's phenotype and a ring chromosome. *J. Clin. Endocrinol. Metab.* 44:340–45

34. Zumoff, B., Bradlow, H. E., Finkelstein, J., Boyar, R. M., Hellman, L. 1976. The influence of age and sex on the metabolism of testosterone. *J. Clin. Endocrinol. Metab.* 42:703–6
35. Kelch, R. P., Kaplan, S. L., Grumbach, M. M. 1973. Suppression of the urinary and plasma follicle-stimulating hormone by exogenous estrogens in prepubertal and pubertal children. *J. Clin. Invest.* 52:1122–28
36. Reiter, E. O., Kulin, H. E., Hamwood, S. M. 1974. The absence of positive feedback between estrogen and luteinizing hormone in sexually immature girls. *Pediatr. Res.* 8:740–45
37. Boyar, R. M., Wu, R. H. K., Kapen, S., Hellman, L., Weitzman, E. D., Finkelstein, J. W. 1976. Clinical and laboratory heterogeneity in idiopathic hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* 43:1268–74

Ann. Rev. Med. 1978. 29:521–41
Copyright © 1978 by Annual Reviews Inc. All rights reserved

RENAL TUBULAR ACIDOSIS

◆7296

F. John Gennari, M.D. and Jordan J. Cohen, M.D.
Department of Medicine, Tufts University School of Medicine, and Renal Service,
New England Medical Center Hospital, Boston, Massachusetts 02111

INTRODUCTION

The kidney contributes to the day-to-day stability of acid-base equilibrium by preventing the excretion of filtered bicarbonate and by promoting the excretion of the strong acids generated by metabolic processes. Although the mechanisms normally responsible for bicarbonate reabsorption and acid excretion are not completely understood, these mechanisms may become disproportionately impaired under certain circumstances giving rise to the clinical syndrome known as renal tubular acidosis. It is our purpose here to review the pathogenesis of this syndrome against the backdrop of what is presently known about the renal regulation of acid-base equilibrium.

THE KIDNEY'S ROLE IN REGULATING ACID-BASE EQUILIBRIUM

Micropuncture observations over the past decade have provided convincing evidence that both the reabsorption of bicarbonate and the excretion of acid by the kidney are accomplished by the same fundamental process, namely, the tubular secretion of hydrogen ion (1). Bicarbonate reabsorption is accomplished when secreted hydrogen ions encounter filtered bicarbonate ions and convert them to carbonic acid. Acid excretion occurs when secreted hydrogen ions lower luminal fluid pH, thereby titrating filtered buffers destined for excretion (e.g. phosphate) and promoting the trapping of secreted ammonia. Because each hydrogen ion that is secreted into the lumen, regardless of its fate, is balanced by a bicarbonate ion that is returned to the body, the single process of hydrogen-ion secretion can account for the conservation of existing bicarbonate stores and for the regeneration of bicarbonate consumed during the buffering of endogenous acids.

It is important to bear in mind that at least 98% of the hydrogen ion secreted by the nephron is normally utilized to recapture the relatively large filtered load of bicarbonate. Thus, even a small decrease in overall hydrogen-ion secretion can

0066-4219/78/0401-0521$01.00

produce important bicarbonate losses into the urine. Although less than 2% of the hydrogen ion secreted by the nephron is utilized for acid excretion, this small fraction is also indispensible for the maintenance of body bicarbonate stores; a failure to excrete the daily acid load would result in a progressive decline in body bicarbonate stores even if the kidneys were to retain the capacity to reabsorb filtered bicarbonate normally.

In view of its central importance in the maintenance of normal acid-base equilibrium, the hydrogen transport system within the kidney is examined more closely below.

Tubular Transport of Hydrogen Ion

Because the electrochemical gradient for hydrogen ion opposes its secretion in both the proximal tubule and collecting duct, it is presumed that the secretory movement of this ion is mediated by an active transport process (1). As in the case of all ions transported against concentration and electrical gradients, the net secretion of hydrogen represents the difference between its active transport into the tubular lumen and its passive back-diffusion out of the lumen. Although the relative rate of back-diffusion in this system is not well defined, it is quite possible that alterations in the passive permeability characteristics of a given nephron segment could have an important influence on the amount of hydrogen ion actually available to combine with filtered bicarbonate and/or to be excreted as net acid.

Nephron segments differ considerably in their ability to transport hydrogen ions. These differences appear to be accounted for by the presence or absence of luminal carbonic anhydrase, by the amount of buffer available in the tubular lumen, and by the limiting pH gradient that the segment can sustain. Whether or not differences in the nature of the active transport process itself are also present is unknown.

PROXIMAL TUBULE Because the proximal tubular epithelium poses little barrier to the passive movement of ions and water, it is not surprising that the maximum hydrogen-ion concentration gradient found in this segment is much smaller than that found in the less permeable epithelia that line more distal segments, especially the collecting duct. Limiting gradients of only 0.4–0.7 pH units have been demonstrated in the late proximal tubule of the rat (2), and even smaller limiting gradients are thought to be present in the dog (3). Despite the relatively small concentration difference that can be sustained, large quantities of hydrogen can be secreted by this segment because of the ready availability of filtered bicarbonate and because of the presence of luminal carbonic anhydrase.

Carbonic anhydrase facilitates the conversion to carbon dioxide and water of the carbonic acid formed by the reaction of secreted hydrogen and filtered bicarbonate, thus minimizing the accumulation of carbonic acid in the tubular lumen. The CO_2 thus formed diffuses into the tubular cell where it is rehydrated to form carbonic acid. This acid can then dissociate to provide hydrogen ions to replace those originally secreted into the tubule; the bicarbonate ions generated in equivalent numbers by this process diffuse into the peritubular interstitium.

Thus, the reaction between secreted hydrogen ion and filtered bicarbonate is tantamount to the actual reabsorption of bicarbonate. The importance of this mechanism is reflected in the fact that proximal tubular hydrogen-ion secretion accounts for at least 70–80% of the total hydrogen ion secreted by the kidney.

As noted above, the secretion of hydrogen ions along the length of the proximal tubule creates a small but measurable decrease in luminal pH; thus, buffers other than bicarbonate (e.g. phosphate and ammonia) are also titrated to a small degree. To what extent the hydrogen ions captured by these buffers gain access to the final urine is uncertain because of the possibility that back-titration may occur in the descending limb of Henle's loop (4).

LOOP OF HENLE The extent to which hydrogen-ion secretion normally occurs between the last accessible portion of the proximal tubule and the earliest accessible portion of the distal tubule is unclear. What is known is that no appreciable change in the concentration of bicarbonate in the luminal fluid occurs between these two points (2). Since about 50% of the fluid leaving the proximal convoluted tubule is reabsorbed between these sites (i.e. about 10% of the original filtered load), it follows that about half of the bicarbonate escaping the proximal tubule is also reabsorbed there. Whether this is accomplished in the descending straight portion of the proximal tubule, in the loop of Henle, or in the very earliest portion of the distal tubule is unknown.

DISTAL TUBULE Hydrogen-ion secretion in the distal convoluted tubule is associated with a somewhat larger transepithelial concentration gradient than that observed in the proximal tubule (2). Unlike the gradient in the proximal tubule, however, this hydrogen-ion gradient can be explained almost entirely by the pronounced transepithelial electrical-potential difference (lumen negative) produced by the active transport of other ions (1). Thus, there is no prima facie evidence that active hydrogen-ion transport actually occurs in this segment. By way of further contrast with the proximal tubule, carbonic anhydrase activity is not present at the luminal membrane of the distal tubule, but is confined to an intracellular location (5, 6).

Although acidification of the luminal fluid in this segment titrates ammonia and other nonbicarbonate buffers to a greater degree than in the proximal tubule, it is important to emphasize that the overwhelming bulk of the hydrogen ion secreted by the distal tubule is still directed at the reabsorption of filtered bicarbonate and not at the formation of acid for excretion (2). In fact, under normal conditions, the rate of bicarbonate delivery is by far the most important determinant of the rate of distal tubular hydrogen-ion secretion (2). This is not to say that the capacity to accommodate an increased load of bicarbonate at this site is unlimited. Quite to the contrary, it appears that even modest increases in the distal delivery of bicarbonate result in some additional bicarbonate escaping reabsorption at this site.

COLLECTING DUCT Since bicarbonate does not normally appear in the final urine, it is evident that the small amount of bicarbonate that does escape reabsorption in the distal tubule is normally recaptured in the collecting duct. As the bicarbonate-depleted tubular fluid descends from the cortical distal tubule to the papillary tip, continued hydrogen-ion secretion reduces luminal pH further, titrating filtered nonbicarbonate buffers and facilitating ammonia diffusion and trapping. Thus, most of the hydrogen ions secreted by this segment are directed at the formation of acid for excretion. The terminal collecting duct can sustain a much larger pH gradient than any other nephron segment; the rate of back-diffusion of hydrogen ion does not normally offset its rate of active secretion until the luminal concentration is 1000 times higher than plasma, i.e. until urine pH is approximately 4.5. Sufficient urinary buffer is normally available, however, to allow excretion of the daily acid load at urine pH values well above this normal gradient limitation for collecting-duct hydrogen-ion secretion.

Carbonic anhydrase has recently been shown to be present only in a distinct subpopulation of epithelial cells lining this segment (5, 6). How this striking anatomic segregation of carbonic anhydrase contributes to the functional characteristics of collecting-duct hydrogen-ion secretion is unclear.

Factors Implicated in the Control of Tubular Hydrogen-Ion Transport

Having reviewed some of the characteristics of hydrogen-ion transport along the nephron, we now consider the various factors thought to influence this process (Table 1). Although much of our current understanding of these regulatory influences is based on data derived from acute studies, it is important to bear in mind that those factors that play an important role in the day-to-day regulation of hydrogen-ion secretion, by definition, must be capable of a sustained effect. In the subsequent discussion, therefore, only those influences that might have such a sustained effect are emphasized.

INTERPLAY BETWEEN THE DEMAND FOR SODIUM CONSERVATION AND THE AVAILABILITY OF CHLORIDE FOR REABSORPTION It is axiomatic that, for any given overall rate of sodium reabsorption in the nephron, the relative rates of bicarbonate and chloride reabsorption fix the relationship between the steady-state extracellular concentration of these anions. Moreover, for any given rate of sodium transport, a change in the rate of either chloride or bicarbonate reabsorption would, of necessity, reciprocally alter the rate of reabsorption of the other. Thus, a sustained alteration in hydrogen-ion secretion and in steady-

Table 1 Factors implicated in the control of tubular hydrogen-ion secretion

1. Interplay between the demand for sodium conservation and the availability of chloride for reabsorption
2. Carbon dioxide tension
3. Carbonic-anhydrase activity
4. Tubular ammoniogenesis
5. Aldosterone secretion
6. Parathyroid hormone secretion
7. Body potassium stores

state bicarbonate concentration would occur if there were a persistent disproportion between chloride and sodium transport. In fact, such a disproportion is characteristic of metabolic alkalosis, in which hydrogen-ion secretion (and hence, bicarbonate reabsorption) is markedly increased due to a reduction in the availability of filtered chloride for reabsorption coupled with a continuing demand for sodium conservation (7). Furthermore, the reduction in overall hydrogen-ion secretion characteristic of chronic HCl-acidosis appears to be the consequence of the continued availability of excess filtered chloride for reabsorption[1] (8).

CARBON DIOXIDE TENSION Chronic alterations in arterial carbon dioxide tension result in parallel changes in renal hydrogen-ion secretion and bicarbonate reabsorption (9, 10). Thus, the level of CO_2 tension in the body fluids is an important modulator of overall hydrogen-ion secretion. Although the renal response to sustained changes in arterial CO_2 tension plays an important role in the adaptation to primary respiratory acid-base disturbances, it is not yet clear whether the ambient arterial CO_2 tension is important in the regulation of hydrogen-ion secretion in the day-to-day maintenance of acid-base equilibrium.

CARBONIC ANHYDRASE ACTIVITY As noted earlier, the enzyme carbonic anhydrase facilitates bicarbonate reabsorption in the proximal tubule. The importance of this enzyme is evidenced by the marked bicarbonaturia and metabolic acidosis that results from the administration of acetazolamide, a potent inhibitor of its action (11). In the absence of such pharmacologic inhibition, however, carbonic anhydrase is not rate-limiting for renal hydrogen-ion secretion and thus cannot be assigned a regulatory role under normal circumstances.

TUBULAR AMMONIOGENESIS It has long been known that urinary ammonium, derived from ammonia produced by the renal parenchyma, is of crucial importance in the bulk excretion of acid, given the normal gradient limitation for hydrogen-ion excretion and the limited availability of filtered buffers, (e.g. phosphate). Day-to-day variations in acid excretion are normally associated with parallel changes in ammonium excretion. In addition, the large increase in acid excretion that occurs in response to exogenous-acid loading or to sustained hypercapnia is accounted for almost entirely by an increase in ammonium excretion (8, 9, 12). Despite this close association between changes in hydrogen-ion and ammonium excretion, the precise role of ammonia production in the regulation of acid excretion remains unclear. It has proved extremely difficult to determine whether ammonia production per se is the rate-limiting step regulating hydrogen-ion secretion or whether regulation of ammonia production is secondary to changes in hydrogen-ion secretion produced by other factors (13).

---

[1] Even though acid excretion is increased in HCl-acidosis, it is clear from the sustained reduction in plasma bicarbonate concentration that overall renal hydrogen-ion secretion is markedly decreased.

ALDOSTERONE Aldosterone influences the renal handling of sodium and potassium in the distal nephron, and through these effects may have a secondary influence on hydrogen-ion secretion. Whether aldosterone per se plays a primary role in the regulation of hydrogen-ion secretion, however, is unclear. Evidence favoring this view stems from the demonstration that aldosterone directly influences hydrogen-ion secretion in the isolated turtle bladder (14), and from the observation that total aldosterone deficiency in experimental animals results in metabolic acidosis (15, 16). Evidence to the contrary includes the demonstration that the administration of mineralocorticoids to normals has only a minimal effect on base balance (17, 18), and that the response to exogenous-acid loading is normal in adrenalectomized dogs treated with a fixed dose of replacement mineralocorticoids (19).

PARATHYROID HORMONE Although the acute administration of parathyroid hormone to normal human subjects has been shown to suppress renal hydrogen-ion secretion and to produce transient bicarbonaturia (20), day-to-day fluctuations in parathormone levels do not appear to exert an appreciable effect on acid-base equilibrium. Even sustained elevations in parathyroid hormone levels, such as occur in patients with primary hyperparathyroidism, have no important effect on acid-base values in most instances (21, 22). It is not yet clear what role, if any, the marked secondary elevations in parathyroid hormone play in the reduction in plasma bicarbonate concentration seen in patients with underlying renal disease (23).

BODY POTASSIUM STORES Much has been written about the influence of body potassium stores on the response of the kidney to acute acid-base challenges. However, little information is available to support the notion that changes in body potassium stores per se alter steady-state renal hydrogen-ion transport in man. Neither prolonged potassium loading nor persistent potassium deficiency has a detectable effect on the level of plasma bicarbonate concentration maintained by the kidney (24).

TUBULAR DISORDERS OF HYDROGEN-ION TRANSPORT

Renal tubular acidosis (RTA) is an abnormality of kidney function characterized by persistent hyperchloremic metabolic acidosis in the face of well-preserved glomerular filtration. Because patients with RTA maintain a low plasma bicarbonate concentration and, hence, a low rate of bicarbonate reabsorption, it follows that overall renal hydrogen-ion secretion must be markedly reduced.

Two distinct abnormalities in tubular function that can cause RTA have been identified: (*a*) a reduction in the normal gradient limitation on renal hydrogen-ion secretion, manifested by the inability to acidify the urine normally, and (*b*) a limitation on overall hydrogen-ion secretion, manifested by the excretion of administered bicarbonate at plasma levels lower than normal. We use the term gradient-limitation RTA to describe the first of these abnormalities (synonymous terms include distal RTA, type 1 RTA, and classic RTA) and the term bicarbonate-wasting RTA to describe the second abnormality (synonymous terms include proximal RTA, and type 2 RTA).

Gradient-Limitation Renal Tubular Acidosis

PATHOPHYSIOLOGY *General features* The characteristic feature of gradient-limitation RTA is the inability to lower urine pH maximally, even when plasma bicarbonate concentration is markedly reduced. Normal individuals excrete maximally acid urine (ph 4.5–5.0) in response to only a modest reduction in plasma bicarbonate concentration. By contrast, the urine pH typically fails to fall below 6.0 in gradient-limitation RTA, even in the face of the severe degree of metabolic acidosis (plasma bicarbonate concentration less than 15 meq/l) that often develops (25).

The inability to sustain a large, hydrogen-ion concentration gradient across the collecting-duct epithelium limits the degree to which nonbicarbonate buffers can be titrated. For example, the amount of hydrogen ion excreted in association with phosphate is directly related to the degree of urine acidification; phosphate buffers can accept more than twice as much hydrogen ion at a urine pH of 5.8 than at a urine pH of 6.8 and almost three times as much at a urine pH of 4.8. Similarly, the amount of hydrogen ion excreted as ammonium is directly related to the degree of acidification; more than twice as much hydrogen ion is excreted with ammonium at a urine pH of 5.0 than at a urine pH of 6.5 (25). As a result of the abnormally low gradient limitation on hydrogen-ion secretion in the collecting duct, patients with this form of RTA may fail to excrete sufficient acid to regenerate the bicarbonate consumed in buffering their normal endogenous-acid load.

Despite their persistently low plasma bicarbonate concentration, it is important to recognize that patients with gradient-limitation RTA are readily able to increase overall hydrogen-ion secretion to normal levels when exogenous buffers are administered. This maneuver circumvents the dampening effect of the abnormally low gradient limitation on hydrogen-ion secretion by providing additional proton acceptors in the tubular fluid, thereby delaying achievement of the limiting pH value. For example, the administration of bicarbonate in amounts sufficient to increase plasma levels to normal values results in an immediate, marked increase in hydrogen-ion secretion, and as a result little or none of the administered bicarbonate escapes into the urine (26, 27). Similarly, administration of dibasic phosphate markedly increases hydrogen-ion secretion; in this instance, the increment in hydrogen-ion secretion is reflected initally by an increase in acid excretion and subsequently by a sustained increase in plasma bicarbonate concentration (26, 27).

Nature of the Defect The specific abnormality or abnormalities in the hydrogen-ion transport system that lead to a reduced gradient limitation for hydrogen-ion secretion in the collecting duct have not been fully elucidated. An abnormally low gradient limitation could reflect the presence of either (*a*) a specific defect in the active hydrogen-ion transport system (a "weak pump"), or (*b*) a reduction in the permeability barrier to the passive backflux of hydrogen ion.

Experimental evidence in favor of the weak-pump hypothesis is derived from studies designed to assess collecting-duct hydrogen-ion secretion when the gradient limitation is removed by bicarbonate loading (28). In these studies, patients with gradient-limitation RTA are given sufficient bicarbonate to increase the pH of the urine to a level higher than that of plasma. The transepithelial hydrogen-ion concentration gradient is thereby reversed, allowing passive forces to assist rather than to oppose hydrogen-ion secretion. Under these circumstances, investigators have attempted to assess hydrogen ion secretion in the collecting duct by measuring urinary carbon dioxide tension. The rationale for this assessment is as follows: Because the distal nephron lacks luminal carbonic anhydrase, the carbonic acid generated when hydrogen ions are secreted into a bicarbonate-rich tubular fluid is swept into the urine prior to spontaneous dehydration. The eventual dehydration of this carbonic acid in the renal pelvis and bladder serves to increase urinary CO_2 tension to levels higher than that in the arterial blood. Thus, it is argued, the presence of a high urinary P_{CO_2} during bicarbonate loading can be used as an index of hydrogen-ion secretion in the collecting duct.

In normal individuals subjected to bicarbonate loading, the carbon dioxide tension in the urine exceeds that in the arterial blood by approximately 30 mm Hg (28). Strikingly, in patients with gradient-limitation RTA, there is no appreciable difference between urine and plasma CO_2 tension under these conditions (28). This observation has led to the conclusion that collecting-duct hydrogen-ion transport is reduced or absent (i.e. a weak pump) in these patients even when the gradient limitation is removed.

This conclusion, however, has not been universally accepted. An alternate explanation for this observation has been advanced, namely that the permeability of the collecting duct to carbonic acid is greater in patients with gradient-limitation RTA than in normals; if this were true, the carbonic acid formed during hydrogen-ion secretion would be more rapidly reabsorbed, accounting for the "low" urine P_{CO_2} values observed during bicarbonate loading (29). Present experimental techniques do not permit a clear choice to be made between these two alternative explanations.

Nevertheless, if gradient-limitation RTA proves to be the consequence, at least in some instances, of a weak pump, the question of the specific mechanism underlying this defect still remains. One intriguing possibility is that a localized deficiency of carbonic anhydrase in the collecting duct diminishes the availability of hydrogen ion for secretion. Few data are available that directly assess renal carbonic anhydrase activity in patients with RTA (30), and no attempt has been made to measure carbonic anhydrase activity in specific nephron segments. Recent studies in normal animals and man demonstrated that carbonic anhydrase in the collecting duct is not only normally confined to an intracellular location but is limited to a distinct subpopulation of cells lining this segment (5, 6). This observation should open up a new avenue for investigating the possible role of carbonic anhydrase deficiency in patients with gradient-limitation RTA.

Experimental support for the second theoretical possibility to explain gradient-limitation RTA, namely increased permeability and passive backflux, comes largely from the observation that membrane-damaging drugs can produce this abnormality. For example, amphotericin-B, an antifungal agent, produces potassium wasting and gradient-limitation RTA in man (31). Using the isolated turtle-bladder preparation, investigators have also shown that amphotericin-B markedly impedes secretion of hydrogen ion against an electrochemical gradient but has no effect in the absence of such a gradient (32). Moreover, amphotericin-B causes a large increase in potassium flux in the turtle bladder, as well as smaller increases in the passive fluxes of sodium and chloride (32). Thus, amphotericin-B apparently produces a nonspecific increase in membrane permeability in this tissue. The possibility that a similar, nonspecific permeability defect may be responsible for gradient-limitation RTA in man is an attractive one, since it could easily account for the abnormalities in sodium and potassium transport also seen in this disorder (see below).

Several investigators have attempted to distinguish between the weak-pump and the increased-backflux hypotheses by assessing the response of patients with gradient-limitation RTA to the infusion of sodium sulfate. This maneuver causes urine pH to fall acutely in normal, sodium-avid subjects, presumably because it increases the transepithelial electrical gradient (lumen negative), thereby facilitating hydrogen-ion secretion into the tubule (33). It has been reasoned that a similar fall in urine pH should occur in patients with gradient-limitation RTA, if the defect in net hydrogen-ion transport is due to increased backflux but not if it is due to a weak pump (34, 35). Although the administration of sodium sulfate does result in a fall in urine pH in some patients with gradient-limitation RTA, it does not do so in all (34, 36–39). Furthermore, even when urine pH falls, the abnormal gradient limitation is not completely abolished (i.e. urine pH fails to reach levels as low as 4.5–5.0) except in patients with mild forms of the defect in whom systemic acidosis is not a feature (38, 39).

We believe that these observations provide little insight into the nature of the defect in gradient-limitation RTA. First, the ability of sodium sulfate to increase intraluminal negativity in the collecting duct is dependent upon reabsorption of sodium at this site. Thus, a failure of sodium sulfate infusions to intensify urinary acidification in patients with gradient-limitation RTA could as easily be explained by an impairment in sodium reabsorption as by an impairment in hydrogen-ion secretion. Indeed, as noted below, many patients with gradient-limitation RTA have a diminished capacity to conserve sodium. Secondly, even if urine pH were to fall, the presence of a weak hydrogen-ion pump would not be excluded because such a defective pump might well be capable of transporting more hydrogen ion under conditions of sodium sulfate infusion in which the passive forces favoring secretion might be augmented.

It is apparent from the foregoing that neither a weak pump nor passive backflux has been excluded as a possible explanation for gradient-limitation

RTA. It is not unreasonable to suggest, moreover, that both abnormalities might occur.

ASSOCIATED FEATURES The characteristic defect in hydrogen-ion transport in patients with gradient-limitation RTA is almost invariably accompanied by an inability to conserve sodium and potassium normally. As a result, untreated patients with this defect often become severely hypokalemic and/or frankly volume depleted. Strikingly, correction of the acidosis by the administration of bicarbonate results in significant repair of both sodium and potassium deficits (27, 40) and in many instances plasma potassium values return to normal levels. The mechanism by which bicarbonate administration ameliorates the sodium and potassium losses, however, is unclear. It is presumed that reabsorption of the administered bicarbonate facilitates sodium transport in the proximal nephron and this permits repletion of extracellular volume. Aldosterone secretion, which is markedly increased in untreated patients with this disorder, falls towards normal after correction of the acidosis and repletion of sodium stores (27). This reduction in aldosterone secretion may contribute to the observed decrease in potassium excretion, but cannot be held totally responsible for it; when sodium chloride rather than sodium bicarbonate is used to reexpand extracellular fluid volume in these patients, there is a similar reduction in aldosterone secretion but no decrease in potassium excretion (27). It seems reasonable to suggest that the fall in potassium excretion during sodium bicarbonate therapy reflects, at least in part, a reduction in distal delivery of sodium.

Although the potassium- and sodium-wasting tendencies are improved when the acidosis is corrected, there is evidence to suggest that these defects in renal function are not always totally reversed (41). For example, despite restoration of normal plasma bicarbonate levels, some degree of hypokalemia often persists unless sizable potassium supplements are also given. Similarly, some patients with this form of RTA fail to conserve sodium normally when challenged with a sodium-restricted intake, even when acid-base equilibrium is restored to normal (42).

Another characteristic feature of gradient-limitation RTA, often even more troublesome than the inability to conserve sodium and potassium, is the demineralization of bone and the tendency to develop nephrocalcinosis. The removal of calcium from bone and the resultant hypercalciuria are almost certainly secondary to the buffering of retained acid by the alkaline calcium salts residing in osseous tissue. A limited number of balance observations in patients with gradient-limitation RTA suggest that daily acid retention occurs even though plasma bicarbonate concentration remains stable, albeit at a low level (43). This observation implies that some nonextracellular buffer mechanism is called into play under these circumstances; the most likely candidate is the large store of calcium carbonate contained in bone. The remarkable healing of bony lesions during sustained correction of the acidosis is consistent with this concept. The frequency with which nephrocalcinosis develops can be accounted for

Table 2 Etiologic classification of gradient-limitation RTA

- Primary (idiopathic)
  - Familial
  - Sporadic
- Drug or toxin induced
  - Amphotericin B
  - Lithium
  - Toluene
- Miscellaneous causes
  - Sjogren's syndrome
  - Lupoid hepatitis
  - Isolated hypergammaglobulinemia
  - Hepatic cirrhosis
  - Transplant rejection
  - Chronic hydronephrosis
  - Nephrocalcinosis

by the concurrence of hypercalciuria, the relatively alkaline urine pH, and the low urinary citrate levels seen in this disorder; all of these conditions favor the intrarenal precipitation of calcium salts.[2]

ETIOLOGY It is convenient to divide the causes of gradient-limitation RTA into three broad categories (Table 2). The primary (idiopathic) form of this disease can occur as an inherited disorder but very often appears sporadically. In addition, as mentioned earlier, this disorder can be produced by certain drugs or toxins. Finally, a number of pathologic states have been clearly associated with gradient-limitation RTA. The reader is referred to several recent reviews for a more detailed discussion of the broad spectrum of diseases in which this disorder is seen (29, 45, 46).

DIAGNOSIS The diagnosis of gradient-limitation RTA depends upon the recognition both of the characteristic changes in plasma composition and of the inability to acidify the urine. Patients with gradient-limitation RTA typically present with a relatively severe hyperchloremic acidosis (plasma bicarbonate concentration often lower than 15 meq/l), and marked hypokalemia (potassium concentration frequently less than 3.0 meq/l), without a persistent elevation in serum creatinine. In such patients, the finding of a urine pH greater than 6.0 virtually establishes the diagnosis; in the face of metabolic acidosis, the absence of a more acid urine indicates that an abnormal gradient limitation must be

[2] It should be noted that some patients have been described in whom nephrocalcinosis clearly precedes the development of systemic acidosis (38, 44). In one instance, hypercalciuria was present for at least three years prior to the development of overt nephrocalcinosis and an acidification defect (44). These observations have led to the postulate that, in certain individuals, deposit of calcium in the medullary portion of the kidney is the cause, rather than the effect, or gradient-limitation RTA (44).

present. Bone disease and medullary nephrocalcinosis, when present, are helpful in supporting the diagnosis.

In acidotic patients with less severe reductions in plasma bicarbonate concentration (16–21 meq/l), the abnormal gradient limitation is usually less severe as well. Regardless of the severity of the acidosis, however, the finding of a random urine pH less than 5.5 serves to exclude the diagnosis. It is useful to recall that urine pH is normally at or near its diurnal nadir just before breakfast; thus, a urine sample obtained at this time may be of greatest diagnostic value (39). In patients in whom the diagnosis of gradient-limitation RTA is under consideration but in whom the diurnal pattern of urine pH never encompasses a level lower than 5.5, a carefully performed ammonium chloride–loading test can provide useful diagnostic information (25, 45). If sufficient ammonium chloride is administered to reduce plasma bicarbonate to a level of approximately 20 meq/l or lower, and if urine pH falls below 5.5, the diagnosis of gradient-limitation RTA can be excluded. Ordinarily, a dose of 0.1 g NH_4Cl/kg-body-weight is adequate for this purpose.

The use of the standard ammonium chloride–loading test has uncovered a group of individuals who do not have spontaneous acidosis but who are nevertheless unable to decrease urine pH below 5.5. Because of this abnormal response, the term incomplete RTA has been applied to such individuals. It is presumed that these people have an abnormal gradient limitation but that it is not sufficiently severe to prevent complete excretion of their normal endogenous-acid load. Incomplete RTA has been noted in the following settings: (*a*) in close relatives of patients with unequivocal gradient-limitation RTA (44), (*b*) in patients with a variety of diseases known to be associated with gradient-limitation RTA (Table 2), and (*c*) in rare patients with recurrent nephrolithiasis and nephrocalcinosis (38). On the basis of this last association, some have advocated that ammonium chloride–loading tests be performed in all patients with unexplained nephrolithiasis in the hope of uncovering a treatable disease (45). However, the time and effort required to identify the rare stone-forming patient with an acidification defect would appear to outweigh any putative therapeutic value to be gained. It appears to us that the only clear indication for acid-loading tests in patients with normal acid-base equilibrium is to identify relatives of patients with familial, gradient-limitation RTA who may have an increased risk of developing nephrocalcinosis and acidosis in the future.

THERAPY Alkali therapy is almost always indicated in patients with gradient-limitation RTA, both to arrest the destructive bone changes and to diminish the potentially life-threatening potassium wasting. The administration of as little as 1 meq/kg per day may be sufficient to correct the acidosis; it is unusual for the alkali requirement to exceed 2–3 meq/kg per day.

The potassium deficiency in patients with gradient-limitation RTA may produce such severe muscle weakness that respiratory effort is seriously impaired. This occasionally results in an acute medical catastrophe. When respiratory failure superimposes an element of respiratory acidosis on the underlying metabolic acidosis, plasma pH may fall to potentially fatal levels. In such patients, urgent correction of the acidosis and of the potassium deficiency must proceed simultaneously; the shift of potassium into cells during rapid correction of acidosis, could, in the absence of concomitant potassium replacement, acutely exacerbate the hypokalemia and further embarrass ventilation. After repair of the often sizeable potassium deficits, continued potassium supplementation is usually required in order to offset any lingering tendency of the kidney to waste this cation.

Bicarbonate-Wasting Renal Tubular Acidosis

PATHOPHYSIOLOGY *General features* The characteristic feature of bicarbonate-wasting RTA is a limitation in overall hydrogen-ion secretion by the kidney, a defect manifested by the failure to retain administered bicarbonate despite the presence of a persistently low plasma bicarbonate concentration. In contrast to gradient-limitation RTA, in which the abnormality appears to be confined to the collecting duct, precise localization of the defect in hydrogen-ion secretion in bicarbonate-wasting RTA is less certain. It is presumed that the proximal tubule is at least partly at fault, because the rate of bicarbonate excretion induced by alkali loading can exceed the rate at which bicarbonate normally escapes proximal reabsorption (47). Indeed, it is possible that the defect involves only the proximal tubule. Given that the distal nephron has only a limited capacity to recapture large bicarbonate loads, it is not necessary to invoke any impairment of distal hydrogen-ion secretion in order to account for the observed rates of bicarbonate excretion.

Although bicarbonate wasting can be demonstrated during the administration of exogenous bicarbonate, it is important to emphasize that patients with this disorder do not excrete abnormal quantities of alkali spontaneously; the absence of persistent bicarbonaturia can be inferred from the observation that plasma bicarbonate concentration, although reduced below normal, remains stable rather than falling continuously. Under such steady-state conditions, the diurnal pattern of urinary acidification can be presumed to be normal. Moreover, a further reduction in plasma bicarbonate concentration readily results in a lowering of urine pH to the same minimal values achieved by normals (48). Such observations show clearly that hydrogen-ion secretion in the terminal collecting duct is unimpaired in this disorder.

Given that the ability to acidify the urine is normal, it is likely that filtered phosphate and secreted ammonia are titrated sufficiently to allow acid excretion to keep pace with the daily endogenous-acid load. Thus, in contrast to gradient-limitation RTA, retention of hydrogen ion and dissolution of bone salts need not occur. The observation that patients with long-standing, but isolated, bicarbonate-wasting RTA fail to develop bone disease or hypercalciuria suggests that normal external hydrogen-ion balance can be maintained in this disorder (49, 50). Indeed, formal acid-balance studies in two patients are available to document this expectation (50).

Nature of the defect A defect in hydrogen-ion secretion involving the proximal tubule could theoretically occur as the result of (*a*) an abnormality in luminal carbonic anhydrase, (*b*) an abnormality in the regulation of proximal hydrogen-ion secretion, or (*c*) an abnormality in the mechanism of hydrogen-ion transport per se. Given the wide variety of circumstances in which bicarbonate-wasting RTA occurs (Table 3), it would not be surprising if clinical examples of each of these abnormalities were found. Indeed, there is evidence favoring precisely this view.

Although carbonic anhydrase deficiency appears not to be the cause in most cases of bicarbonate-wasting RTA (48), it has been implicated in at least one family with this disorder (51). In this kindred, three members with severe bicarbonate wasting were found to have a mutant carbonic anhydrase with extremely low activity in red blood cells. Although renal tissue was not available for similar study, it was inferred that the same enzymatic defect was responsible for the renal tubular abnormality.

Of the regulatory factors thought to influence proximal tubular hydrogen-ion secretion (Table 1), only parathormone has been postulated to play a pathogenetic role in bicarbonate-wasting RTA (46). High parathormone levels, which are thought to dampen proximal bicarbonate reabsorption, are in fact a characteristic feature of the many patients who have bicarbonate-wasting RTA in association with elements of the Fanconi syndrome, i.e. those who have hyperphosphaturia and hypocalcemia (see below). Moreover, correction of the hypocalcemia in these patients results in a reduction in parathormone levels and a marked improvement in bicarbonate reabsorption (46). Even more impressive

Table 3 Etiologic classification of bicarbonate-wasting RTA

- Primary (idiopathic)
  - Familial
    - Isolated bicarbonate wasting
    - Associated with Fanconi syndrome
  - Sporadic
    - Isolated bicarbonate wasting
    - Associated with Fanconi syndrome
- Drug or toxin induced
  - Acetazolamide
  - Outdated tetracycline
  - Streptozotocin
- Miscellaneous causes
  - Cystinosis
  - Lowe's syndrome
  - Hereditary fructose intolerance
  - Tyrosinemia
  - Light-chain proteinuria
  - Multiple myeloma
  - Transplant rejection

evidence favoring a role for parathormone has been obtained from a single patient with hypoparathyroidism and coincidental hereditary fructose intolerance, a rare disorder in which bicarbonate-wasting RTA is induced by the ingestion of fructose. This patient failed to manifest bicarbonate wasting while ingesting fructose until exogenous parathormone was administered (52). Although these observations indicate that parathormone may play an important role in some patients with bicarbonate-wasting RTA, it should be recalled that clinically significant acidosis is not a feature of isolated (i.e. primary) hyperparathyroidism (21, 22). Therefore, additional abnormalities almost certainly contribute to the bicarbonate wasting in patients with this form of RTA even when parathormone levels are elevated.

Little attention has been paid to the possibility that a malfunction of other putative regulatory factors is responsible for some cases of bicarbonate-wasting RTA. For example, a primary impairment of proximal tubular sodium transport or a primary enhancement of chloride reabsorption could, in theory, lead to the observed reduction in proximal tubular hydrogen-ion secretion. One could even speculate that a reduced sensitivity of the proximal tubule to the effects of carbon dioxide tension on bicarbonate reabsorption might result in bicarbonate wasting.

The remaining possible cause of bicarbonate-wasting RTA is a faulty hydrogen-ion transport system. Indeed, this would seem to be the only plausible explanation in the absence of a defect in carbonic anhydrase or of a malfunction of any of the factors that regulate hydrogen-ion transport. This possibility is supported indirectly by the presence of other defects in proximal tubular transport (e.g. impaired glucose, phosphate, amino acid, and/or uric acid reabsorption) in most patients with bicarbonate-wasting RTA. Because of the likely proximal location of the defect in bicarbonate-wasting RTA, it is obvious that evaluation of the adequacy of the hydrogen-ion transport mechanism itself will prove even more difficult than in the case of gradient-limitation RTA. Consequently, the precise mechanism(s) responsible for bicarbonate-wasting RTA are likely to remain elusive.

ASSOCIATED FEATURES In most patients with bicarbonate-wasting RTA, the impairment of hydrogen-ion secretion is accompanied by other abnormalities of proximal tubular function, including defects in glucose, phosphate, uric acid, and amino acid reabsorption (Fanconi syndrome). In some cases, however, as noted above, the hydrogen-ion secretory defect is the only detectable abnormality in tubular transport.

Unlike the case with gradient-limitation RTA, plasma potassium is often normal or only slightly reduced in the bicarbonate-wasting form of this disorder (53). By way of further contrast, attempts to correct the acidosis by the administration of bicarbonate, a maneuver that typically ameliorates potassium wasting in gradient-limitation RTA, characteristically causes plasma potassium concentration in patients with bicarbonate wasting to fall, often to dangerously low levels (40, 41). This finding can easily be explained, since, as noted above, the administration of sodium bicarbonate in the presence of a defect in proximal reabsorption results in the delivery of increased amounts of these ions to the distal nephron, thereby providing a potent stimulus for distal potassium secretion (54).

ETIOLOGY As shown in Table 3, there are two forms of primary (idiopathic) bicarbonate-wasting RTA. The first is characterized by an isolated defect in hydrogen-ion secretion, and has been noted to occur almost exclusively in infants. This disorder typically remits spontaneously within several years (55). The second is characterized by associated defects in proximal tubular function (Fanconi syndrome) and usually persists indefinitely. Both types of primary bicarbonate-wasting RTA can occur sporadically or in a familial pattern. In addition, a number of drugs, toxins, and disease states have been found to produce this type of RTA, almost always in association with Fanconi syndrome. The reader is referred to several recent reviews for a more detailed discussion of the etiology of this disorder (29, 45, 46).

DIAGNOSIS A firm diagnosis of bicarbonate-wasting RTA depends on the recognition of persistent hyperchloremic acidosis and of an inability to retain administered bicarbonate. However, it is rarely necessary to carry out a formal bicarbonate-loading test; readily available laboratory tests usually suffice to distinguish this form of RTA from the gradient-limitation form. For example, since patients with bicarbonate-wasting RTA have no acidification defect, a urine sample obtained before breakfast (when urine pH is normally acid) should have a pH of less than 5.5. In addition, since most adult as well as many younger patients with bicarbonate-wasting RTA have one or more additional abnormalities in proximal tubular function, the findings of glycosuria, aminoaciduria, hypophosphatemia, and/or hypouricemia give strong support to the diagnosis. Those patients with isolated bicarbonate wasting may be more difficult to distinguish from patients with gradient-limitation RTA, but even in such individuals routine laboratory studies are helpful; the finding of a normal plasma potassium concentration (before treatment) in a patient with hyperchloremic acidosis favors the presence of a bicarbonate-wasting defect rather than a gradient-limitation defect.

Finally, the response to a therapeutic trial of alkali can aid in the separation of the two forms of RTA. If large quantities of bicarbonate (i.e. greater than 3–4 meq/kg per day) are required to sustain a normal plasma concentration, it can be concluded that the kidney is unable to conserve bicarbonate normally. If, in addition, bicarbonate therapy aggravates rather than ameliorates potassium wasting, the diagnosis of bicarbonate wasting is virtually certain.

THERAPY In contrast to the clear benefit of bicarbonate therapy in gradient-limitation RTA, the advantages to be gained from alkali treatment in bicarbonate-wasting RTA are less certain. Part of the problem is that, when bicarbonate is administered, it is usually excreted so rapidly into the urine that little improvement occurs unless very large quantities are used. Moreover, as noted earlier, alkali administration in these patients often results in the emergence of serious hypokalemia. Fortunately, in patients with only a mild depression of plasma bicarbonate concentration, the acidosis per se appears to have no important morbidity, and can probably be safely ignored. There are some circumstances, however, in which the inconvenience and potential hazards of alkali therapy are clearly justified. For example, in infants with an isolated defect in hydrogen-ion secretion, alkali administration appears to be necessary to insure normal growth and development (55). Also, in the unusual instance in which plasma bicarbonate is severely depressed, some alkali replacement may be essential to provide a margin of protection against an additional, superimposed cause of acidosis.

From the observation in normal individuals that overall hydrogen-ion secretion can be increased by stimulating proximal sodium reabsorption, it has been reasoned that bicarbonate reabsorption might also be augmented in patients with bicarbonate-wasting RTA by stimulating tubular sodium conservation. Indeed, persistent, mild volume depletion produced by daily hydrochlorothiazide administration was successful in treating several patients with bicarbonate-wasting RTA (56–58). Unfortunately however, this treatment does not restore plasma bicarbonate concentration completely to normal levels unless combined with the administration of moderate amounts of bicarbonate (56, 57). In addition, potassium wasting is markedly exacerbated; serum potassium often falls to levels as low as 2–2.5 meq/l unless very large potassium supplements are also given (56, 57). This unexplained complication clearly limits the usefulness of diuretic therapy to patients who truly require treatment but who cannot tolerate the administration of large quantities of alkali.

Hyperchloremic Acidosis of Renal Insufficiency

It is well known that metabolic acidosis is virtually an invariable feature of end-stage renal disease; as renal damage becomes progressively more severe, a point is reached at which hydrogen-ion secretion in the remaining functioning tubules is simply inadequate to keep pace with the ordinary demands for hydrogen-ion excretion. Under these circumstances, plasma unmeasured-anion concentration[3] is also typically increased. Some patients with only mild impairment of renal function, however, develop a metabolic acidosis without an increase in unmeasured-anion concentration, i.e. an acidosis that is hyperchloremic in nature. Under these circumstances, plasma acid-base composition is indistinguishable from the pattern seen in renal tubular acidosis.

Although first noted in patients with predominant tubulointerstitial disease (59), it has become apparent that this hyperchloremic acidosis is a reasonably common feature of mild renal insufficiency. Because the ability to maximally acidify the urine is unimpaired in these patients, it is evident that their persistent acidosis is not the consequence of a gradient-limitation defect in hydrogen-ion transport. These considerations imply that a bicarbonate-wasting tendency is responsible and, indeed, this contention is supported by the few direct observations that are available (29, 60). There is no clear explanation as yet for the emergence of this defect in bicarbonate conservation in patients with mild renal insufficiency. In fact, it is not even possible to localize the site of impaired bicarbonate reabsorption with certainty, since in these individuals the amount of bicarbonate that escapes reabsorption during bicarbonate loading does not typically exceed the amount normally reabsorbed in the distal nephron.

---

[3] Plasma unmeasured anion concentration is defined as the plasma sodium concentration minus the sum of the plasma bicarbonate and chloride concentrations.

This observation coupled with the low levels of aldosterone often found in these patients led to the speculation that aldosterone deficiency could be responsible for the observed acidosis by retarding distal bicarbonate reabsorption (29, 60, 61). This possibility is supported by the observation that the acidosis is ameliorated by the administration of mineralocorticoid in some (62) but not in all instances (60, 61).

Other possibilities that might be advanced to explain the persistent hyperchloremic acidosis in mild renal insufficiency include (*a*) dampened proximal bicarbonate reabsorption due to elevated parathyroid hormone levels or to other hormonal abnormalities, (*b*) enhancement in chloride reabsorption due to some metabolic alteration, and (*c*) direct tubular damage due to the underlying renal disease. The data presently available do not allow one to choose from among these possibilities.

CONCLUSIONS

The present understanding of the renal mechanisms responsible for the regulation of acid-base equilibrium in normal individuals offers a basis for the differential diagnosis and treatment of patients with renal tubular acidosis. Recent experimental findings have led to the recognition that many different pathogenetic mechanisms can produce each of the clinical patterns of renal tubular acidosis. What remains for future investigation is the precise delineation of the cellular and molecular basis for the transport defects that underlie this interesting group of disorders.

ACKNOWLEDGMENTS

This work was supported in part by Grants HL-00759, HL-15810, from the National Heart and Lung Institutes, and AM-18446 from the National Institute of Arthritis, Metabolism, and Digestive Diseases.

Literature Cited

1. Rector, F. C. Jr. 1976. Renal acidification and ammonia production; chemistry of weak acids and bases; buffer mechanisms. In *The Kidney*, ed. B. M. Brenner, F. C. Rector Jr., pp. 318–43. Philadelphia: Saunders, 1948 pp.
2. Malnic, G., de Mello Aires, M., Giebisch, G. 1972. Micropuncture study of renal tubular hydrogen ion transport in the rat. *Am. J. Physiol.* 222:147–58
3. Clapp, J. R., Owen, E. E., Robinson, R. R. 1965. Contribution of the

proximal tubule to urinary ammonia excretion by the dog. *Am. J. Physiol.* 209:269–72
4. Glabman, S., Klose, R. M. Giebisch, G. 1963. Micropuncture study of ammonia excretion in the rat. *Am. J. Physiol.* 205:127–32
5. Rosen, S. 1972. Localization of carbonic anhydrase activity in the vertebrate nephron. *Histochem. J.* 4:35–48
6. Lonnerholm, G. 1973. Histochemical demonstration of carbonic anhydrase activity in the human kidney. *Acta Physiol. Scand.* 88:455–68
7. Schwartz, W. B., van Ypersele de Strihou, C., Kassirer, J. P. 1968. Role of anions in metabolic alkalosis and potassium deficiency. *New Engl. J. Med.* 279:630–39
8. deSousa, R. C., Harrington, J. T., Ricanati, E. S., Shelkrot, J. W., Schwartz, W. B. 1974. Renal regulation of acid-base equilibrium during chronic administration of mineral acid. *J. Clin. Invest.* 53:465–76
9. Schwartz, W. B., Brackett, N. C. Jr., Cohen, J. J. 1965. The response of extracellular hydrogen ion concentration to graded degrees of chronic hypercapnia: the physiologic limits of the defense of pH. *J. Clin. Invest.* 44:291–301
10. Gennari, F. J., Goldstein, M. B., Schwartz, W. B. 1972. The nature of the renal adaptation to chronic hypocapnia. *J. Clin. Invest.* 51:1722–30
11. Maren, T. H. 1967. Carbonic anhydrase: Chemistry, physiology, and inhibition. *Physiol. Rev.* 47:595–781
12. Sartorius, O. W., Roemmelt, R. C., Pitts, R. F. 1949. The renal regulation of acid-base equilibrium in man. IV. The nature of the renal compensations in ammonium chloride acidosis. *J. Clin. Invest.* 28:423–39
13. Schwartz, W. B., Cohen, J. J. 1978. On the nature of the renal response to chronic disorders of acid-base equilibrium. *Am. J. Med.* In press
14. Al-Awqati, Q., Norby, L. H., Mueller, A., Steinmetz, P. R. 1976. Characteristics of stimulation of H^+ transport by aldosterone in turtle urinary bladder. *J. Clin. Invest.* 58:351–58
15. Kurtzman, N. A., White, M. G., Rogers, P. W. 1971. Aldosterone deficiency and renal bicarbonate reabsorption. *J. Lab. Clin. Med.* 77:931–40
16. Hulter, H. N., Ilnicki, L. P., Harbottle, J. A., Sebastian, A. 1977. Impaired renal H^+ secretion and NH_3 production in mineralocorticoid-deficient glucocorticoid-replete dogs. *Am. J. Physiol.* 232:F136–46
17. Relman, A. S., Schwartz, W. B. 1952. The effect of DOCA on electrolyte balance in normal man and its relation to sodium chloride intake. *Yale J. Biol. Med.* 24: 540–58
18. Kassirer, J. P., London, A. M., Goldman, D. M., Schwartz, W. B. 1970. On the pathogenesis of metabolic alkalosis in hyperaldosteronism. *Am. J. Med.* 49:306–15
19. Cohen, J. J., Hulter, H. N., Smithline, N., Melby, J. C., Schwartz, W. B. 1976. The critical role of the adrenal gland in the renal regulation of acid-base equilibrium during chronic hypotonic expansion. *J. Clin. Invest.* 58:1201–8
20. Hellman, D. E., Au, W. Y. W., Bartter, F. C. 1965. Evidence for a direct effect of parathyroid hormone on urinary acidification. *Am. J. Physiol.* 209:643–50
21. Coe, F. L. 1974. Magnitude of metabolic acidosis in primary hyperparathyroidism. *Arch. Intern. Med.* 134:262–65
22. Mallette, L. E., Bilezikian, J. P., Heath, D. A., Aurbach, G. D. 1974. Primary hyperparathyroidism: clinical and biochemical features. *Medicine* 53:127–46
23. Muldowney, F. P., Donohoe, J. F., Carroll, D. V., Powell, D., Feaney, R. 1972. Parathyroid acidosis in uraemia. *Q. J. Med.* 41:321–42
24. Tannen, R. L. 1977. Relationship of renal ammonia production and potassium homeostasis. *Kidney Int.* 11:453–65
25. Wrong, O., Davies, H. E. F. 1959. The excretion of acid in renal disease. *Q. J. Med.* 28:259–313
26. Reynolds, T. B. 1958. Observations on the pathogenesis of renal tubular acidosis. *Am. J. Med.* 25:503–15
27. Gill, J. R. Jr., Bell, N. H., Bartter, F. C. 1967. Impaired conservation of sodium and potassium in renal tubular acidosis and its correction by buffer anions. *Clin. Sci.* 33:577–92
28. Halperin, M. J., Goldstein, M. B., Haig, A., Johnson, M. D., Stine-

baugh, B. J. 1974. Studies on the pathogenesis of Type I (distal) renal tubular acidosis as revealed by the urinary $PC0_2$ tensions. *J. Clin. Invest.* 53:669–77

29. Sebastian, A., McSherry, E., Morris, R. C. Jr. 1976. Metabolic acidosis with special reference to the renal acidoses. *See Ref. 1,* pp. 615–60
30. Yaffe, S. J., Craig, J. M., Fellers, F. X. 1960. Studies on renal enzymes in a patient with renal tubular acidosis. *Am. J. Med.* 29:168–75
31. McCurdy, D. K., Frederic, M., Elkinton, J. R. 1968. Renal tubular acidosis due to amphotericin B. *N. Engl. J. Med.* 278:124–31
32. Steinmetz, P. R., Lawson, L. R. 1970. Defect in urinary acidification induced in vitro by amphotericin B. *J. Clin. Invest.* 49:596–601
33. Schwartz, W. B., Jenson, R. L., Relman, A. S. 1955. Acidification of the urine and increased ammonium excretion without change in acid-base equilibrium: Sodium reabsorption as a stimulus to the acidifying process. *J. Clin. Invest.* 34: 673–80
34. Gyory, A. Z., Edwards, K. D. G. 1971. Effect of mersalyl, ethacrynic acid and sodium sulfate infusion on urinary acidification in hereditary renal tubular acidosis. *Med. J. Aust.* 2:940–45
35. Nascimento, L., Rademacher, D. R., Hamburger, R., Arruda, J. A. L., Kurtzman, N. A. 1977. On the mechanism of lithium-induced renal tubular acidosis. *J. Lab. Clin. Med.* 89:455–62
36. Seldin, D. W., Wilson, J. D. 1966. Renal tubular acidosis. In *The Metabolic Basis of Inherited Diseases.* ed. S. W. Stanbury, J. B. Wyngaarden, D. S. Frederickson. pp. 1230–46. New York: McGraw-Hill, 2nd edition. 1434 pp.
37. Morris, R. C., Piel, C. F., Audioun, E. 1965. Effects of sodium phosphate and sulfate on renal acidification in two patients with renal tubular acidosis. Pediatrics 36:899–904
38. Buckalew, V. M. Jr., McCurdy, D. K., Ludwig, G. D., Chaykin, L. B., Elkinton, J. R. 1968. Incomplete renal tubular acidosis: physiologic studies in three patients with a defect in lowering urine pH. *Am. J. Med.* 45:32–42
39. Tannen, R. L., Falls, W. F. Jr., Brackett, N. C. Jr. 1975. Incomplete renal tubular acidosis: some clinical and physiological features. *Nephron* 15:111–23
40. Sebastian, A., McSherry, E., Morris, R. C. Jr. 1971. On the mechanism of renal potassium wasting in renal tubular acidosis associated with the Fanconi Syndrome (Type 2 RTA). *J. Clin. Invest.* 50:231–43
41. Sebastian, A., McSherry, E., Morris, R. C. Jr. 1971. Renal potassium wasting in renal tubular acidosis (RTA). Its occurrence in types 1 and 2 RTA despite sustained correction of systemic acidosis. *J. Clin. Invest.* 50:667–78
42. Sebastian, A., McSherry, E., Morris, R. C. Jr. 1976. Impaired renal conservation of sodium and chloride during sustained correction of systemic acidosis in patients with Type 1, classic renal tubular acidosis. *J. Clin. Invest.* 58:454–69
43. Goodman, A. D., Lemann, J. Jr., Lennon, E. J., Relman, A. S. 1965. Production, excretion and net balance of fixed acid in patients with renal acidosis. *J. Clin. Invest.* 44: 495–506
44. Buckalew, V. M. Jr., Purvis, M. L., Shulman, M. G., Herndon, C. N., Rudman, D. 1974. Hereditary renal tubular acidosis. *Medicine* 53:229–54
45. Narins, R., Goldberg, M. 1977. Renal tubular acidosis: pathophysiology, diagnosis and treatment. *Dis. Mon.* 23:1–66
46. Morris, R. C. Jr., Sebastian, A., McSherry, E. 1972. Renal acidosis. *Kidney Int.* 1:322–40
47. Morris, R. C. Jr. 1969. Renal tubular acidosis: Mechanisms, classification and implications. *N. Engl. J. Med.* 281:1405–13
48. Rodriguez-Soriano, J., Edelmann, C. M. Jr. 1969. Renal tubular acidosis. *Ann. Rev. Med.* 20:363–82
49. Brenes, L. G., Brenes, J. N., Rodriguez, V. M., Hernandez, M. M. 1977. Familial proximal renal tubular acidosis. A distinct clinical entity. *Am. J. Med.* 63:244–52
50. Lemann, J. Jr., Wilz, D. R., Brenes, L. G. 1976. Acid, calcium and phosphorus balances in proximal renal tubular acidosis. *Kidney Int.* 10:561 (Abstr.)
51. Shapira, E., Ben-Yoseph, Y., Eyal, F. G. Russell, A. 1974. Enzymat-

ically inactive red cell carbonic anhydrase B in a family with renal tubular acidosis. *J. Clin. Invest.* 53: 59–63

52. Morris, R. C. Jr., McSherry, E., Sebastian, A. 1971. Modulation of experimental renal dysfunction of hereditary fructose intolerance by circulating parathyroid hormone. *Proc. Natl. Acad. Sci. USA.* 68: 132–35
53. Lightwood, R., Payne, W. W., Black, J. A. 1953. Infantile renal acidosis. *Pediatrics* 12:628–44
54. Malnic, G., de Mello Aires, M., Giebisch, G. 1971. Potassium transport across renal distal tubules during acid-base disturbances. *Am. J. Physiol.* 221: 1192–1208
55. Nash, M. A., Torrado, A. D., Greifer, I., Spitzer, A., Edelmann, C. M. Jr. 1972. Renal tubular acidosis in infants and children. *J. Pediatr.* 80:738–48
56. Rampini, S., Fanconi, A., Illig, R., Prader, A. 1968. Effect of hydrochlorothiazide on proximal renal tubular acidosis in a patient with idiopathic "deToni-Debre-Fanconi syndrome." *Helv. Paediatr. Acta* 23: 13–23
57. Donckerwolcke, R. A., van Stekelenburg, G. J., Tiddens, H. A. 1970. Therapy of bicarbonate-losing renal tubular acidosis. *Arch. Dis. Child.* 45:774–79
58. Callis, L., Castello, F., Fortuny, G., Vallo, A., Ballabriga, A. 1970. Effect of hydrochlorothiazide on rickets and on renal tubular acidosis in two patients with cystinosis. *Helv. Paediatr. Acta* 25:602–19
59. Lathem, W. 1958. Hyperchloremic acidosis in pyelonephritis. *N. Engl. J. Med.* 258: 1031–36
60. Perez, G. O., Oster, J. R., Vaamonde, C. A. 1974. Renal acidosis and renal potassium handling in selective hypoaldosteronism. *Am. J. Med.* 57:809–16
61. Vagnucci, A. H. 1970. Selective aldosterone deficiency in chronic pyelonephritis. *Nephron* 7:524–37
62. Sebastian, A., Schambelan, M., Lindenfeld, S., Morris, R. C. 1977. Amelioration of metabolic acidosis with fludrocortisone therapy in hyporeninemic hypoaldosteronism. *N. Engl. J. Med.* 297:576–83

Ann. Rev. Med. 1978. 29:543–62
Copyright © 1978 by Annual Reviews Inc. All rights reserved

BEHAVIORAL RISK FACTORS IN CORONARY ARTERY DISEASE

◆7297

C. David Jenkins, Ph.D.
Department of Behavioral Epidemiology, Boston University School of Medicine,
Boston, Massachusetts 02118

The belief that personality, emotions, and disturbing life experiences play a part in the development of cardiovascular diseases is widely held, both by the general public and by the medical practitioners who treat their illnesses. This possible association is treated with great skepticism, however, by many of those academic and research physicians and biologists who are more laboratory or statistically oriented. Consensus between practitioners and researchers is slow in developing, not because the clinical observations lack merit, but because of the great difficulty of defining and measuring psychosocial and behavioral concepts. This, in turn, has delayed the rigorous research that is essential to distinguish fortuitous observations from genuine patterns in nature. Recently, more rigorous epidemiological study designs and more quantitative techniques for assessing psychological and social factors have permitted substantial scientific progress. The role of psychosocial and behavioral factors in cardiovascular disease, particularly in coronary artery disease (CAD), has already become important to the teacher of medicine, and may soon take on sufficiently valid operational forms as to be of utility to the medical practitioner.

This review focuses first on that segment of social-behavioral research into the etiology of CAD that we believe to be currently the most advanced—that dealing with the "coronary-prone behavior pattern" (Type A). We then review additional psychosocial risk factors whose association with CAD has been shown repeatedly, and finally, close with discussion about implications for medical education and practice.

HISTORICAL BACKGROUND

The basic idea of an association of emotions and cardiovascular function goes back historically much further than most people realize. Three hundred fifty years ago a young physician with new and wild ideas wrote the following: "Every affection of the mind that is attended with either pain or pleasure, hope or fear, is the cause of an agitation whose influence extends to the heart." (1) The

0066-4219/78/0401-0543$01.00

year was 1628; the Latin monograph in which this appeared was *Exercitatio De Motu Cordis et Sanguinis in Animalibus*, and the author's name was Dr. William Harvey.

While William Harvey's observation applied to the general functioning of the heart, a famous British surgeon 150 years later saw the connection between his own experience with angina pectoris and his emotions and behavior. Dr. John Hunter (1729–93) once stated, "My life is in the hands of any rascal who chooses to annoy me. . . ." Dr. Hunter died suddenly while attending a board meeting at St. George's Hospital in London in 1793.

A hundred years later, another of the immortals of medicine, Sir William Osler, was among those to observe that the then quite rare condition of atherosclerosis and its attending episodes of angina pectoris were not spread randomly throughout the community, but rather tended to attack a specific type of person. Osler wrote in 1897 ". . . I believe that the high pressure at which men live, and the habit of working the machine to its maximum capacity are responsible for (arterial degeneration) rather than excesses in eating and drinking. . . ." (2). Other early examples could be cited supporting the basic hypothesis that emotions and behavior relate to the presence of coronary disease.

In our century, Osler's description of the coronary patient striving for success in a responsible but high-pressured way was elaborated upon by the famous American psychiatrists, the Menningers (3), who saw in coronary patients strong aggressive tendencies, which were often strictly repressed. Kemple (4) affirmed this pattern of aggressivenss in his observations of coronary disease patients and added that they were usually very ambitious and strove compulsively to achieve goals incorporating power and prestige.

The Coronary-Prone Behavior Pattern

Perhaps the greatest progress toward an integrated description of the psychological and behavioral contributions to CAD has come from the on-going work of the cardiologists Meyer Friedman and Ray H. Rosenman. For twenty years they have pursued a series of laboratory, clinical, and epidemiological studies of the role of behavior and the central nervous system in the development of CAD. Friedman conceptualizes the psychosocial contribution to CAD risk as the coronary-prone behavior pattern or Type A behavior. He defines this behavior pattern as "a characteristic action-emotion complex which is exhibited by those individuals who are engaged in a relatively *chronic struggle* to obtain an *unlimited* number of *poorly-defined* things from their environment in *the shortest period of time*, and, if necessary, against the opposing efforts of other things or persons in the same environment." (5) In his latest writings, Friedman asserts that the most critical aspects of the Type A behavior pattern are excesses of aggression, hurry, and competitiveness, all of which are acted out as a struggle to overcome barriers in the environment.

People who manifest this behavior pattern are called Type A individuals, and those who have the opposite style of behavior—who are relaxed, unhurried, mellow, and more easily satisfied— are designated Type B. The Type B person is also interested in progress and achievement, in most instances, but more often tends to flow with the stream of life rather than constantly struggle against it as the Type A person does.

Persons having the Type A pattern are often so deeply committed to their vocation or profession that other aspects of their lives are relatively neglected. Not all features of this pattern need be present for a person to be classified as Type A. The pattern is not an abstract psychodynamic inference nor is it an automatic reaction to an environmental situation, but rather it is the reaction of a psychologically predisposed person to a situation that challenges him or her. Different kinds of situations provoke maximal reactions from different persons, depending on their values and needs, but most predisposed individuals exhibit a consistency over time.

It is important to distinguish Type A behavior from the ill-defined concept of stress. The term stress is used in a variety of ways, sometimes to refer to a painful stimulus or upsetting situation, and at other times to refer to a personal reaction of alarm, discomfort, or pain. The divergent ways in which the notion of stress has been treated in the hands of many researchers in this field has thwarted the progress of psychosomatic research and has raised the level of skepticism among biological scientists toward all social and psychological research into disease etiology. In contrast to stress, however defined, the Type A behavior pattern is neither a stressor situation nor a distressed response. It is rather a style of overt behavior with which some people confront life situations, either pleasant or troubling, provided that some element of challenge is felt to be present.

The most widely used measures of the coronary-prone behavior pattern to date have been the Structured Interview of Rosenman, Friedman, et al (6) and the Activity Survey of Jenkins et al (7). The Structured Interview consists of a set of 25–30 questions and observations made of the study subject. Behavior typing is based more on the manner in which the subject responds to the interviewer than on the content of responses. The several editions of the self-administered Activity Survey questionnaire consist of from 54 to 64 multiple-choice questions scored by computer according to algorithms developed by cross-validation against assessments based on the Structured Interview.

EVALUATION CRITERIA

In order to provide a rubric for judging the scientific merit of the hypothesis that the Type A behavior pattern and other behavioral risk factors are etiologically associated with CAD, we refer to a set of criteria by which epidemiologists often judge whether the evidence from an aggregate of studies supports a presumed cause-effect relationship. We ask the following questions of the various studies purporting to relate behavioral risk factors to CAD:

1. What is the strength of the association between the predictor variable and the emergence of disease?
2. Is this a consistent finding? Has it been replicated by other workers in other places?

3. Have studies established that the predictor actually precedes the outcome condition?
4. Is the predictor or risk factor associated rather specifically with one outcome (in this case, CAD), or is it associated with many different disease conditions?
5. Can a biological gradient be demonstrated? That is, does a more intense presence of the risk factor lead to a higher probability of disease or more severe disease?
6. Is the new finding coherent with earlier knowledge about the antecedent and outcome variables?
7. Are the implications of the observed findings biologically sensible?
8. Is there experimental evidence, either in humans or in animals, in which the biological outcome has been produced under controlled administration of the psychosocial predictor condition?

Each of these criteria are applied below to the research evidence that exists on coronary-prone behavior.

Strength of Association

Several cross-sectional studies rated behavior type by means of the Structured Interview, and performed diagnostic procedures for the presence of clinical CAD, each variable being gathered without knowledge of the other. Like most epidemiological studies of CAD, these were largely limited to male study groups because of the much higher incidence of CAD in that sex. The relative-risk ratios derived by dividing CAD prevalence rates for Type A persons by those for Type B persons are relatively consistent. Analysis of intake data from the Western Collaborative Group Study (WCGS) reveals that Type A men have 2.2 times the prevalence of CAD compared to Type B men (6). In the North American Monastery Study by Quinlan, Caffrey, and colleagues (8), Type A monks had 2.3 times the prevalence of angina and 4.3 times the prevalence of infarction as compared to Type B monks.

A more recent study of persons entering the Chicago Heart Association Screening in Industry Project (9) utilized the Activity Survey questionnaire rather than the Standard Interview as the means for measuring Type A behavior. Examination of the Type A score without reference to other factors yielded a nonsignificant association with the presence of CAD. When other risk factors, particularly age, were statistically controlled, however, the Type A score became a statistically significant correlate of coronary disease prevalence, which had a standardized regression coefficient about the same as those for serum cholesterol and diastolic blood pressure. This implies it would generate a similar relative-risk ratio to these other risk factors in this population, probably on the order of 2.0 or 2.5 to 1. The Screening in Industry Study has the added advantage of utilizing a multivariate procedure that isolates the independent contributions of risk factors. In this study, the contribution of Type A behavior to risk prediction was found to be independent of other risk factors.

Prospective studies are less subject to biases than cross-sectional studies and lend further support to the Type A hypothesis. At several times during the 8½-year WCGS, analyses of the relative risk of developing new CAD were performed, both for the Standard Interview and for the self-administered Activity Survey. The relative-risk ratios generated differed by age of the cohort and length of follow-up. They are summarized in Table 1. The table shows a steady relative-risk ratio for men 50–59 years old, with Type A men having about twice the new CAD as Type B men. For men 39–49 years old, the relative-risk ratio declines over time. The reason for this is not well understood, but it may derive in various degrees from (*a*) rapid emergence of CAD in the most coronary-prone of the younger cohort, which would leave a less susceptible residue in later years; (*b*) greater instability of the behavior pattern in the younger than older group, which would result in increasing amounts of misclassification due to changed risk factors in later years; and (*c*) other unknown phenomena. The phenomena are not the results of selective dropout from the Study inasmuch as this was very small (12).

The multiple-logistic model was used to determine the relative risk of CAD incidence attributable solely to Type A behavior after other risk factors were statistically controlled (13). Only minimal diminution of the observed relative-risk ratios resulted, which suggests that the Type A behavior contribution to CAD risk is essentially independent of the risk conferred by more standard biological measures.

The one prospective analysis of effectiveness of the Activity Survey in predicting CAD incidence covered the last four years of the WCGS. Men with the highest Type A scores, the top third of the population distribution, incurred 1.7 times the incidence of new CAD as the tercile of men most in the Type B direction (14). Multivariate analyses comparing the relative strengths of the Standard Interview and the Activity Survey revealed the Standard Interview to be the stronger predictor of new CAD. The associations of both measures with CAD were essentially independent of other major risk factors. Combined use of the Standard Interview and the Activity Survey gave somewhat greater strength of prediction than either procedure alone (15).

In summary, studies thus far published show that the Type A behavior pattern, as currently measured, has about the same strength of association with CAD prevalence and incidence as do other standard risk factors. Type A men seem to be at about twice the risk of CAD compared with Type B men. Higher relative-risk ratios are found when the population at risk is younger, and the

Table 1 Relative risk of CAD incidence: Type A/Type B

	Length of follow-up		
Age of cohort at intake (yr)	2½ yr (10)	4½ yr (11)	8½ yr (12)
39–49	6.5	2.7	1.9
50–59	1.9	1.7	2.1

assessment of the behavior pattern is no more than two to three years distant. In several studies where a variety of risk factors were compared, the Type A pattern was shown to be independent of all other standard risk factors in its contribution to the prediction of CAD.

Replication and Consistency of Findings

A surprising number of clinical and epidemiological studies of CAD confirm the association of the Type A behavior pattern, or particular traits of that pattern, with prevalence or incidence of CAD. Jenkins' 1971 review paper (16) cited 15 studies in Europe and the United States published by researchers other than the San Francisco group (Friedman & Rosenman) and the Boston group (Jenkins & Zyzanski), which showed that one or more aspects of the Type A behavior pattern measured in various ways were related to CAD in cross-sectional studies. Such elements as "striving aggressively toward achievement," "perfectionistic," "aggressive," "unable to relax," and "tendency to overwork" were implicated. Among these studies, two teams used the structured behavior type interview and found an association between Type A behavior and coronary disease in both hospital patients (17, 18) and residents of monasteries (19, 8). Between 1971 and 1975, considerably more research was performed. A 1976 review paper lists 24 studies dealing with Type A or related behavioral phenomena (20). These studies varied in strength of research design, but all had adequate sample sizes. Twenty-two of these studies showed positive associations with CAD, one had equivocal results, and one had negative findings. In addition to the WCGS, prospective studies of achievement and job involvement were reported from Belgium, a Swedish study associated time pressures and excess job responsibility with CAD, and one other study involving interviews for the Type A pattern revealed it to have positive associations with CAD (21).

The validity of the Activity Survey questionnaire as a correlate of coronary disease is being shown in a lengthening series of retrospective studies in different settings. Following the initial retrospective finding that the Type A scale of the 1965 Activity Survey discriminated CAD cases as a group in the WCGS (7), later editions of the Survey were found valid in other settings—with hospitalized patients in Bridgeport, Connecticut (22) and with Veterans Administration patients in Durham, North Carolina (23), Houston, Texas (24), and St. Petersburg, Florida (D. N. Hiland, unpublished data). The type A pattern was associated with CAD prevalence in the Honolulu Heart Program, but only in one of the two decades studied and only when combined with either the presence of social mobility or status incongruity (25). In 1976, a Polish translation of the Activity Survey significantly distinguished between groups of CAD patients and healthy controls (K. Wrzesniewski, unpublished data).

In summary, the consistency of the findings relating Type A behavior pattern —or selected elements thereof—to CAD is quite remarkable. Certainly this record is stronger than that compiled for such "second-level" risk factors as physical exercise, diet, or elevated serum levels of glucose or uric acid. The international findings, coming as they do from the Soviet Union, Poland, the

Netherlands, Belgium, Israel, Sweden, and Australia indicate the cross-cultural relevance of at least selected aspects of the Type A pattern and also their probable contribution to CAD risk in many kinds of communities and cultures.

Time Sequence

Can it be established that the Type A behavior pattern actually precedes coronary disease? Can it be argued that Type A behavior is merely the expression of subclinical changes in the health of cerebral and cardiac vasculature? The several prospective studies described above present important evidence that Type A behavior in fact precedes the clinical emergence of CAD. The prospective study of Swedish twins by Floderus (26) is an important addition to the prospective WCGS studies. The research of Bonami & Rime (27) argues strongly that a striving for achievement and an intensive dedication to career are indeed quite likely to precede all but the earlier stages of atherosclerosis. The projective tests on which this study was based were administered an average of 9.7 yr before the emergence of clinical CAD. Friedman et al (28) make the important point that many psychosocial precursors of CAD may in fact be expressions of subclinical disease. This argument seems most plausible for such variables as anxiety, depression, somatic complaints, and disturbed sleep. It seems less plausible for characteristics of the Type A pattern such as increased robustness, activity, and expenditure of energy. The hypothesis, nevertheless, deserves further examination.

Other evidence that the Type A behavior pattern actually precedes disease comes from studies of the early development of this behavior pattern. Bortner et al (29) were able to distinguish between Type A and Type B behaviors in adolescent boys. More recent work shows that this distinction is possible in elementary school children (K. Matthews, unpublished data). Repeated-measures studies show the Type A pattern to be a stable personal characteristic quite resistant to change. All these data suggest that Type A behavior—or its converse, Type B—become built-in as enduring characteristics sometime during childhood at a time generally considered to precede formation of the earliest arterial plaques.

Specificity of the Risk Factor

Although evidence is sketchy, the current state of knowledge suggests that the Type A pattern is specifically associated with atherosclerotic diseases. Studies of mortality in the WCGS show that only those persons dying of coronary disease and those dying of accidents have a clearly Type A mean score on the Activity Survey. (Those dying of accidents also have elevated Speed and Impatience scores.) Men dying of malignancies, lung diseases, and all other categories tend to be rather evenly divided between Type A and Type B behaviors, with occasional and perhaps insignificant excesses at times in the Type B category (C. D. Jenkins, unpublished data).

The Southeastern Connecticut Study (30) found coronary disease patients to be more Type A than other hospitalized patients or normal controls. However, incidence of Type A behavior in the other hospitalized patients was not significantly higher than that in the controls.

A hospital study showed a tendency for patients with cerebrovascular accidents (CVA) to score slightly more in the Type A direction on the Activity Survey than patients hospitalized for a variety of problems, excluding cardiac and vascular diagnoses (31). When the CVA patients were further divided into those with and without prior history of atherosclerotic or coronary heart disease, it was found that the group with such a prior history had substantially higher Type A scores than CVA patients without this cardiovascular complication, who scored at about the average of the healthy population.

Further evidence for the specificity of the Type A pattern comes from studies of psychological factors in patients undergoing coronary angiography. Zyzanski et al in Boston (32) found among 95 men who underwent coronary angiography for clinical reasons that those who had two or more coronary vessels obstructed 50% or more by atherosclerosis scored, on the average, very strongly in the Type A direction on all four scales of the Activity Survey. In contrast, the remaining patients who were found to have less severe atherosclerosis scored substantially in the Type B direction.

This study also spoke to another question of specificity: Is the behavior pattern related to degree of arterial obstruction independently of the degree of symptomatology experienced? A two-way analysis of covariance applied to the data showed that the relation of Type A score to degree of arterial obstruction remained significant after the reported intensity of angina pectoris, presence or absence of myocardial infarction, and age of the patient were all taken into account statistically (32).

The general thrust of these studies is that the Type A behavior pattern is rather specifically associated with CAD and conditions involving atherosclerosis and is not found with more-than-average frequency in patients with other types of disease. This finding needs further testing in patients with still other categories of disease, particularly classical psychosomatic disorders.

Dose-Response Relationship Between Behavior Type and CAD

An association between a presumed risk factor and a disease outcome is less likely to be due to artifact if a biological gradient of association can be demonstrated—that is, if a more intense presence of a risk factor is associated with a higher probability of the disease or more severe manifestation of the disease. The method by which the Standard Interview for the Type A behavior pattern is currently quantified precludes examination of this issue. The Interview simply judges persons dichotomously as Type A or Type B. The Activity Survey, however, provides a continuous score, which permits us to address this issue, even though the Activity Survey is probably a less sensitive measure of the behavior pattern than the Standard Interview. Two studies to date provide evidence that a continuous positive relationship exists between strength of Type A and probability of CAD. The incidence study using the Activity Survey (14) showed a stepwise relation between Type A score and incidence of CAD; the top, middle, and lowest terciles in score had successively descending incidence rates. A study of recurrent myocardial infarction shows that persons who have incurred or are about to incur their second infarction have higher mean scores on the Activity Survey than those who have only one infarction; the latter group scores higher, in turn, than persons without CAD in any form (33). In this sample, the Type A score was the strongest predictor of recurrent infarction of the six risk factors included in the analysis. Even after forcing the entry of four standard risk factors (cigarette usage, serum cholesterol, diastolic blood pressure, and age) into the multiple-discriminant function prediction equation, Type A score retained an undiminished statistical significance. Both of these studies support the presence of a biological gradient of association between Type A behavior and CAD.

Coherence with Earlier Knowledge

There is a relatively short history of quantitative research with the Type A behavior pattern, hence it is difficult to assess the coherence of recent work with earlier psychological research. Type A measures correlate rather poorly (usually less than $r = .25$) with most psychological constructs used by clinical and social psychologists or psychiatrists. Recently, however, Glass and his colleagues (24) offered interesting theoretical integrations of Type A behavior with such social psychology concepts as learned helplessness, the need to control, and the striving for power.

Attempts to relate Type A behavior to the mainstream of knowledge in physiology and medicine are just beginning. This is a much needed research area. Present evidence permits the initiation of discussion of the physiologic and pathogenic processes by which Type A behavior is translated into CAD (5, 34).

It is widely argued that the existing standard risk factors for CAD explain only a part of its etiologic picture, and that additional elements await discovery to improve both the prediction and biological explanation of this disease. It is our judgment that behavioral factors, particularly Type A, can contribute substantially to the further definition of CAD etiology.

Biological Plausibility

An important test of new knowledge is its plausibility in the light of the total framework of existing scientific information. The Type A hypothesis presumes a major role of the central nervous system in cardiovascular regulation and in the potential for creating disordered function and eventually pathological structural changes. The known impact of behavior and psychological states on blood pressure, serum free fatty acids, serum lipids, platelet aggregation, blood-clotting mechanisms, endocrine activity, and sympathetic nervous system activity reaching the heart and major vessels argues strongly that any behavioral predisposition to intense, frequent, and sustained central nervous system arousal (such as may well occur in the intensely active Type A person or in persons suffering chronic anxiety, depression, neurotic conflict, or sleep disturbance) increases the probability of disordered cardiovascular function, and eventually, ensuing disease. It is not known why some of these behavioral manifestations (e.g. Type A behavior) happen to have a predilection for affecting the heart, while other behavioral expressions seem equally able to affect a variety of organ systems. A number of excellent reviews and books examined the specifics of central nervous system involvements in cardiovascular function (35–37).

Summing Up Type A Behavior Findings

Considering all eight categories of evidence above, one may reasonably conclude that there is a well-established association between the kinds of behavior referred to in the coronary-prone behavior pattern and CAD prevalence, incidence, recurrence, and underlying atherosclerotic pathology. The findings are unlikely to be spurious and unlikely to be secondary to any of the biological or demographical variables traditionally used in epidemiologic studies of CAD. The findings are not in conflict with earlier knowledge about atherogenesis, and indeed may provide a point of departure for increasing scientific knowledge in this general area.

ADDITIONAL PSYCHOSOCIAL RISK FACTORS

Although, at present, research into Type A behavior speaks more comprehensively to the eight methodological questions that introduced this review, Type A behavior by no means represents the full contribution of psychosocial factors to the prediction and explanation of CAD. Type A behavior closely parallels William Osler's understanding of behavior and CAD, but it seems only to reflect a part of the implications of William Harvey's statement and even less of John Hunter's. Over the past thirty years, a broad array of social and psychological variables has been associated with CAD in clinical and epidemiological studies. A number of comprehensive reviews of this body of research were published before 1970 (41–45), and additional reviews have appeared covering more recent years (16, 46, 20). The following sections comment only on those variables that have received the most replication as correlates of CAD.

Indices of Social Position and Social Change

Epidemiologic research, relying upon sociological and demographic indicators considered singly, has often yielded significant findings, but the direction of these findings has not been consistent from one nation to the next, or even within the same nation over different decades of observation. It would seem that gross social indicators are too far removed from the critical biological processes that give rise to CAD.

A more promising approach to social and demographic variables has been their use in combinations and configurations to generate such constructs as social mobility and status incongruity. Sociologists perceive that the social equilibrium and internal value systems of persons affected by these conditions are exposed to greater instability.

Social mobility refers to changes in persons' environment—both changes occurring between generations and those developing in the lifetime of the people concerned. Status incongruity is the condition of possessing simultaneously the identifying markings of different social classes, as evidenced by discrepancies in levels of education, occupation, income, quality of housing, and organizational membership. Status incongruity is thought to generate conflicting messages, which create uncertainty in a person's self-concept and inconsistency in the way one is treated by others.

A number of studies showed a relation between social mobility and CAD prevalence; it was found in California (47), in Pennsylvania (48), in Chicago (49), in the North American Monastery Study (19), and in North Dakota (50). This last study found that the influence of social mobility on CAD was independent of smoking history, reported hypertension, relative body weight, dietary fat, and parental longevity. Other workers looked with equal diligence and did not find an association between social mobility and coronary disease. This includes the work of Hinkle and associates (51, 52), who studied 270,000 employees of the Bell System. Negative findings were also reported from Connecticut (30). Furthermore, the strong studies of Israeli government employees conducted by Medalie et al (53) did not provide clearcut support for the prospective relationship of social mobility to myocardial infarction. However, geographic mobility (i.e. having been a resident of three or more countries) was associated with a higher future incidence of new angina pectoris (54).

The strongest evidence favoring the relation of status incongruity to CAD emerged from the carefully controlled prospective study of the Western Electric Company in Chicago (49). Cross-sectional studies conducted in Pennsylvania (48, 55) and in Oklahoma (56) supported such a relation, as well.

Further limited support for a relation between these variables and CAD comes from the cross-sectional analysis of social data gathered in the Honolulu Heart Program. There, in one of the two age groups studied, CAD prevalence was significantly higher in men who possessed the combination of Type A score (from the Activity Survey) with either social mobility or status incongruity (25).

Although both supporting and detracting evidence has emerged in recent years for the social mobility–status incongruity hypothesis, the overall pattern of evidence suggests that these two factors may be valid predictors of CAD risk only in certain places and eras, only for certain presentations of CAD (e.g. angina pectoris), or only when certain other variables are also present (i.e. an interaction effect). These variables deserve further examination in carefully controlled prospective studies to resolve these questions.

Anxiety and Neuroticism

Anxiety and neuroticism are central to psychiatric practice and research. In addition, there are a number of quite reliable, valid, and economical methods by which to measure tendencies toward neurotic conflict or the presence of painful affects. A comprehensive review in 1971 (16) found 26 published reports dealing with psychological precursors of CAD. A review that followed in 1976 reported

19 additional studies (20). These studies were highly variable, both in their methods and findings.

It can readily be seen that retrospective studies attempting to relate anxiety and neuroticism to CAD are fraught with problems of selective survival and retrospective bias. Recent prospective studies overcome these vulnerabilities. Taking both retrospective and prospective research into account, the general thrust of the findings is that this family of variables is associated with CAD.

There are more inconsistencies in this segment of the literature than in the studies of the coronary-prone behavior pattern, but the anxiety and neuroticism studies used many different methods of assessing emotional states and varied greatly in adequacy of research design. In general, these variables are more consistently related to angina pectoris than they are to myocardial infarction. In fact, the only associations of anxiety and neuroticism to myocardial infarction emerged from retrospective studies. The association of these variables with angina pectoris was reported from both prospective and retrospective studies, which provides greater confidence in the validity of this specific association. The anxiety-neuroticism dimension represents one more risk factor that operated differently for angina than for infarction.

Recent studies asked questions about sleep disturbance, including insomnia and early awakening (20). These six studies, three prospective and three retrospective, all showed positive associations between sleep difficulties and CAD. Sleep difficulties, more than many other psychological phenomena, imply substantial central nervous system involvement.

In contrast to sociologic indices, measures of emotional conflicts and painful affects reach rather directly into the day-by-day rhythms of life, complete with their discordant overtones. Hence, it would seem that as one moves across the continuum of psychosocial variables toward those variables that implicate the central nervous system, one observes much more consistent associations with CAD. The findings for the coronary-prone behavior pattern are consistent with this generalization.

Life Dissatisfactions

Life dissatisfaction, a combination of both personal psychology and environmental stress, includes complaints about life circumstances as well as the report of problems of the social environment. The bulk of studies dealing with these variables are retrospective. They repeatedly found associations between report of dissatisfactions and myocardial infarction and angina pectoris. Many of these findings could simply be the result of the greater tendency of patients with CAD or any other serious disease to be more critical and complaining.

Recently, two prospective studies advanced our knowledge regarding these variables. In a large prospective study of Swedish twins, Floderus (26) inquired about family problems, work dissatisfactions, and behavioral indicators of "psychosocial discord." Higher levels of reported discord were associated with later emergence of angina pectoris (by report on the London School of Hygiene

Questionnaire) and with early deaths from "All Causes." Data for myocardial infarction were not available. The 10,000-person Israeli study by Medalie et al (54) showed that problems and conflicts in the area of finance, family, work, coworkers, and superiors were all associated prospectively with more than a double risk for the development of angina pectoris. The same items and subscales, however, were not associated with elevated risk of new myocardial infarction in the same population (53).

This family of variables appears to relate to CAD in much the same way as anxiety and neuroticism: higher levels of problems are reported after the clinical emergence of CAD, both by angina and infarction patients. However, in the prospective sense, these variables are predictive only of future development of angina pectoris but not of future myocardial infarction.

To the best of our knowledge, none of these studies of anxiety, neuroticism, or life dissatisfaction, except for the Kaiser-Permanente study, which did not find positive associations (28), incorporated adequate statistical control for other standard risk factors to determine the independence of the contribution to CAD risk of the behavioral risk factor.

Stress and Life Change

While anecdotal reports abound in relation to this category of variables, there are, in addition, a number of adequately controlled studies that support the notion that painful or sustained environmental insults (ranging from death of close relatives to prolonged work overload) can raise the risk of angina pectoris, myocardial infarction, and coronary death (16).

More recently, a substantial number of reports have appeared relating amount of life change, particularly as measured by the Holmes and Rahe Schedule of Recent Experience, to the presence of CAD. Retrospective reports include a study of 39 cardiac deaths in Stockholm (57), a study of 54 survivors of myocardial infarction in the same city (58), a project involving 279 survivors of infarction in Helsinki (59), and another study of 30 outpatients in Sweden (60). All four of these research projects used patients (or surviving next-of-kin) as their own control group, comparing their reports of life changes immediately preceding their acute CAD event with earlier periods in their life as recalled at the same interview. These studies found increased life-change scores [calculated from the Holmes & Rahe scale (73)] reported for the six months immediately preceding infarction. The scores were characteristically two to three times the magnitude of those obtained for prior six-month periods in the life history. The patterns of scores differed, however, between persons for which this was the first episode of cardiovascular illness and those with a prior history of illness. The retrospective design of these studies coupled with the lack of a control group leaves these findings vulnerable to criticism on several methodological grounds.

A retrospective study comparing Swedish women with myocardial infarction to a healthy reference group, conducted by a different research team (61), found a higher reported frequency of troubling social events in the immediate past in the infarction patients. This finding tends to support the life-change research reported above.

Lundberg, et al (62) used the Holmes and Rahe Schedule of Recent Experience with infarction patients and a matched control group in a retrospective study. These workers found no statistical differences between the two groups when total life-change units were calculated by the standard method, but they found significant differences when each person assigned to the events experienced scores reflecting his own perception of the impact of events. Methodologically, the strongest test of the life-change hypothesis to date is the prospective study of 6579 Swedish construction workers ages 41–61 yr by Theorell and colleagues (63). They found no association between elevated life-change scores and incidence of acute myocardial infarction over the ensuing 12–15 mo. Elevated scores did, however, predict development of neuroses and certain other chronic disabling conditions.

In a population-based study involving systematic sampling of a cohort of men in East and West Finland, Rahe et al (64) found East Finns to have substantially higher life-change scores on the average than the West Finns who were sampled. This difference parallels the long-observed excess of coronary disease in East Finland compared to other parts of this nation. This study has the strength of representative sampling, but the weakness of representing only an ecologic association. Continuing surveillance of the two population samples will in time provide additional prospective data, which will assist in judging the validity of the hypothesis that increased intensity of life changes may represent a risk factor for CAD.

METHODOLOGICAL EVALUATION OF ADDITIONAL PSYCHOSOCIAL CATEGORIES

Research into the CAD risk implications of sociologic variables, anxiety and neuroticism, life dissatisfactions, and stress and life change is presently less comprehensive than the research into the coronary-prone behavior pattern. For this reason, less can be said concerning the methodological criteria for deciding the general validity and causal implications of the observed associations. We therefore attempt to summarize the state of knowledge with regard to each of the listed methodological questions for all these categories collectively.

Studies of sociologic indicators and CAD have been mostly crosssectional or retrospective, except for the Western Electric and Israeli studies. (we are not including studies based on national census and mortality data.) As stated earlier, there is a disconcerting lack of consistency among studies performed in nations at different stages of the transition from an agrarian to an industrialized society. In general, relative-risk ratios were not generated, and no evidence was presented for a biological gradient. A variety of biological mechanisms have been offered as links between the sociologic observation and the medical endpoint, but the evidence is inadequate to argue strongly for any one such biological linkage. However, those studies that controlled statistically for biological risk factors found the social factor to be independently associated with CAD, thus ruling out risk-altering behaviors such as dietary practices and cigarette smoking as likely mediating links (65, 50).

Inasmuch as most studies of psychosocial factors have been case-control or cross-sectional studies of nonrepresentative samples, the measure of association has been in terms of statistical significance to "rule out chance differences," rather than in terms of relative-risk ratios. The prospective Israeli studies, however, report relative-risk ratios for their brief anxiety scale and indicators of life dissatisfaction in the range of 2.0–2.5. This is the same magnitude reported for the coronary-prone behavior pattern and for many standard biological risk factors (depending on the thresholds used for separating high-risk from low-risk persons).

Serious consideration must be given to the overall evidence that neuroticism, anxiety, and painful affects relate to risk of CAD. The sheer number of such positive findings from many social settings, despite the wide variety of psychological measurement approaches used, argues that this is a robust association, particularly in terms of angina pectoris. The gist of the studies is that the time-sequence relationship is prospective for angina pectoris but not for myocardial infarction. Both of these clinical conditions, however, raise the levels of anxiety, depression, and emotional problems, thus affecting retrospective studies. The differential elevations of these problems in the period immediately after the acute episode may have important implications for prognosis, however (66). The work of Floderus (26) suggests that the association of these affects is not specific to CAD but is common to a variety of clinical conditions, including psychiatric disturbances and mortality from "All Causes."

The findings with regard to anxiety and neuroticism are coherent with earlier knowledge and biologically plausible in that these affects directly involve the central nervous system, and much laboratory research with humans and animals shows the direct and profound influence of alterations in CNS function on cardiovascular reactivity (44, 67). The prospective and retrospective studies of the relation of sleep disturbance to CAD (20) and the two prospective studies, that relate elevated reactivity to the cold-pressor test with future CAD (68, 69) are further psychophysiological arguments for the plausibility of this category of findings.

A very similar set of comments directed to all these methodological points can be made for variables reflecting life dissatisfactions and interpersonal problems. The reader is referred to two review articles for details (16, 20).

Because most studies of life stress and life-change events have been retrospective, little can be said about strength of association and time sequence. The fact that the one published prospective study in this area found no association cautions us that the retrospective findings may reflect some form of spurious or indirect relationship. The life-change literature suggests that these kinds of stressful or challenging events do not have a specific relationship to CAD, but may raise the risk of a wide variety of medical and psychiatric illnesses. The biological plausibility of this hypothesis rests upon the quite reasonable argument that life changes call upon the organism to expend considerable energy in the process of adaptation and they raise the probability of frequent and sustained departures from emotional and psychophysiological homeostasis. The life-change work is coherent with earlier reports of increased CAD mortality among widows, widowers, and others experiencing a death in the immediate family (70).

There is some evidence for a biological gradient in which greater intensity of the behavioral risk factor is associated with a more severe underlying disease. A study relating anxiety and depression to cineangiographic ratings in living subjects observed that greater degrees of arterial obstruction were associated with higher average anxiety and depression scores and reduced ability to deny unpleasant feelings (32). Additional support comes from an autopsy study wherein next-of-kin reported more severe problems at work among those decedents who had more severely diseased arteries (71). A recent experimental animal study is also in general support of these findings (72). To our knowledge, there have been no large studies of the relation of these several categories of behavioral variables to recurrent infarction.

CURRENT STATUS OF KNOWLEDGE AND IMPLICATIONS FOR PRACTICE

From the data presented in this highly selective review, and from the much broader background of information available in the published literature, the following conclusions are offered:

1. There is ample evidence replicated in many settings that a variety of psychosocial and behavioral factors are consistently associated with the development of coronary disease.
2. The evidence for a few of these categories meets enough of the eight study-design criteria listed and discussed in this paper to suggest that the associations observed are not fortuitous or spurious but have a substantial probability of contributing to etiological understanding of CAD.
3. The structure of evidence is particularly strong for the coronary-prone behavior pattern, Type A. This behavioral variable appears to be antecedent to the clinical emergence of angina pectoris, myocardial infarction, and recurrent myocardial infarction, and seems to be somehow associated with the atherosclerotic process.
4. Although the statistical interaction of behavioral risk factors and biological risk factors in the production of clinical CAD has been analyzed in only a minority of published reports, all such analyses reveal that the behavioral risk factors make an independent contribution to the prediction of CAD risk above and beyond that conferred by the standard biological variables.
5. As is the case for several standard biological risk factors to CAD, there is no evidence at present from natural experiments or clinical field trials whether reducing or eliminating any of these behavioral risk factors will lower the

probability of the clinical emergence of CAD or reduce the rate of its progression.

6. In view of the fact that much still needs to be learned about the clinical value of treating these behavioral risk factors, the clinician who encounters a patient with one or more such factors may choose to pursue one or both of the following courses:

(*a*) Intervene aggressively to reduce such biological risk factors as may be elevated. This is based on the recognition that the biological and behavioral risk factors are probably additive and on the presumption that there is better evidence that biological risk factor reduction has a higher probability of CAD prevention.

(*b*) Work to reduce or eliminate those behavioral risk factors where such a change is justified on grounds that it would improve the quality of life without creating hardship or sacrifice for the patient. Working to reduce the intensity and chronicity of anxiety, depression, or an intense life dissatisfaction would fall into this category, as would providing guidance to increase the ability to cope with life changes. Counseling or encouraging self-directed modification of environmental constraints or personal habits so as to reduce the "hurry sickness," aggressiveness, and behavioral intensity of persons with the Type A behavior pattern might well improve interpersonal harmony and the enjoyment of life, in addition to whatever benefits it might confer in terms of reduced CAD risk.

The truism that further research is needed applies most aptly to this area. Meanwhile, however, it is important for medical educators and practitioners to recognize that much progress has been made, and that even now the more solidly substantiated findings can provide a greater depth of social, psychological, and physiological background to the medical practitioner dealing with CAD patients or persons at elevated risk of this disease.

Literature Cited

1. Harvey, W. 1628. *Exercitatio De Motu Cordis et Sanguinis in Animalibus*. Cited in Eastwood, M. R., Trevelyan, H. 1971. Stress and coronary heart disease. *J. Psychosom. Res*. 15:289–92
2. Osler, W. 1897. Lectures on angina pectoris and allied states. Cited in Friedman, M. 1969. *Pathogenesis of Coronary Artery Disease*. New York: McGraw-Hill. 269 pp.
3. Menninger, K. A., Menninger, W. C. 1936. Psychoanalytic observations in cardiac disorders. *Am. Heart J*. 11:10–21
4. Kemple, C. 1945. Rorschach method and psychosomatic diagnosis. Personality traits of patients with rheumatic disease, hypertensive cardiovascular disease, coronary occlusion, and fracture. *Psychosom. Med*. 7:85–89
5. Friedman, M. 1969. See Ref. 2
6. Rosenman, R. H., Friedman, M., Straus, R., Wurm, M., Kositchek, R., Hahn, W., Werthessen, N. T. 1964. A predictive study of coronary heart disease: the Western Collaborative Group Study. *J. Am. Med. Assoc*. 189:15–22
7. Jenkins, C. D., Zyzanski, S. J., Rosenman, R. H. 1971. Progress toward validation of a computer-scored test for the Type A coronary-prone behavior pattern. *Psychosom. Med*. 33:193–202
8. Quinlan, C. B., Barrow, J. G., Hayes, C. G. 1969. *The association*

of risk factors and coronary heart disease in Trappist and Benedictine monks. Presented at Am. Heart Assoc. Conf. Cardiovasc. Epidemiol. 42nd, New Orleans

9. Shekelle, R. B., Schoenberger, J. A., Stamler, J. 1976. Correlates of the JAS Type A behavior pattern score. *J. Chronic Dis*. 29(6):381–94
10. Rosenman, R. H., Friedman, M., Straus, R., Wurm, M., Jenkins, C. D., Messinger, H. B., Kositchek, R., Hahn, W., Werthessen, N. T. 1966. Coronary heart disease in the Western Collaborative Group Study: a follow-up experience of two years. *J. Am. Med. Assoc*. 195:86–92
11. Rosenman, R. H., Friedman, M., Straus, R., Jenkins, C. D., Zyzanski, S. J., Wurm, M. 1970. Coronary heart disease in the Western Collaborative Group Study: a follow-up experience of 4½ years. *J. Chronic. Dis*. 23:173–90
12. Rosenman, R. H., Brand, R. J., Jenkins, C. D., Friedman, M., Straus, R., Wurm, M. 1975. Coronary heart disease in the Western Collaborative Group Study: final follow-up experience of 8½ years. *J. Am. Med. Assoc*. 233:872–77
13. Brand, R. J., Rosenman, R. H., Sholtz, R. I., Friedman, M. 1976. Multivariate prediction of coronary heart disease in the Western Collaborative Group Study compared to the findings of the Framingham Study. *Circulation* 53:348–55
14. Jenkins, C. D., Rosenman, R. H., Zyzanski, S. J. 1974. Prediction of clinical coronary heart disease by a test for coronary-prone behavior pattern. *N. Engl. J. Med*. 290:1271–75
15. Brand, R. J., Rosenman, R. H., Jenkins, C. D., Sholtz, R. I., Zyzanski, S. J. 1977. Comparison of coronary heart disease prediction in the Western Collaborative Group Study using the Structured Interview and the Jenkins Activity Survey assessments of the coronary-prone Type A behavior pattern. Submitted for publication
16. Jenkins, C. D. 1971. Psychologic and social precursors of coronary disease. *N. Engl. J. Med*. 284:244–55, 307–17
17. Keith, R. A., Lown, B., Stare, F. J. 1965. Coronary heart disease and behavior patterns: an examination of method. *Psychosom. Med*. 27:424–34
18. Cassel, J. C. 1966. Letter to the editor. *Psychosom. Med*. 28:283–84
19. Caffrey, B. 1970. A multivariate analysis of sociopsychological factors in monks with myocardial infarctions. *Am. J. Public Health*. 60: 452–58
20. Jenkins, C. D. 1976. Recent evidence supporting psychologic and social risk factors for coronary disease. *N. Engl. J. Med*. 294:987–94, 1033–38
21. Bruhn, J. G., Paredes, A., Adsett, C. A. 1974. Psychological predictors of sudden death in myocardial infarction. *J. Psychosom. Res*. 18: 187–91
22. Kenigsberg, D., Zyzanski, S. J., Jenkins, C. D., Wardwell, W. I., Licciardello, T. 1974. The coronary-prone behavior pattern in hospitalized patients with and without coronary heart disease. *Psychosom. Med*. 36:344–51
23. Stokols, J. J. 1973. *Life dissatisfaction as a risk factor in coronary heart disease*. PhD thesis. Univ. North Carolina, Chapel Hill. 108 pp.
24. Glass, D. C. 1977. Stress, behavior pattern and coronary disease. *Am. Sci*. 65:177–87
25. Cohen, J. B. 1974. *Sociocultural change and behavior patterns in disease etiology: an epidemiologic study of coronary disease among Japanese Americans*. PhD thesis. School of Public Health, Univ. Calif., Berkeley
26. Floderus, B. 1974. Psychosocial factors in relation to coronary heart disease and associated risk factors. *Nord. Hyg. Tidskr. Suppl*. 6
27. Bonami, M., Rime, B. 1972. Approche exploratoire de la personalité pre-coronarienne par analyse standardisée de donneés projectives thematiques. *J. Psychosom. Res*. 16:103–13
28. Friedman, G. D., Ury, H. K., Klatsky, A. L., Siegelaub, M. S. 1974. A psychological questionnaire predictive of myocardial infarction: results from the Kaiser-Permanente Epidemiologic Study of Myocardial Infarction. *Psychosom. Med*. 36: 327–43
29. Bortner, R. W., Rosenman, R. H., Friedman, M. 1970. Familial similarity in pattern A behavior: father

and sons. *J. Chronic Dis.* 23:39–43

30. Wardwell, W. I., Bahnson, C. B. 1973. Behavioral variables and myocardial infarction in the Southeastern Connecticut Heart Study. *J. Chronic Dis.* 26:447–61
31. Gianturco, D. T., Breslin, M. S., Heyman, A., Gentry, W. D., Jenkins, C. D., Kaplan, B. 1974. Personality patterns and life stress in ischemic cerebrovascular disease. I. Psychiatric findings. *Stroke* 5: 453–60
32. Zyzanski, S. J., Jenkins, C. D., Ryan, T. J., Flessas, A., Everist, M. 1976. Psychological correlates of coronary angiographic findings. *Arch. Intern. Med.* 136:1234–37
33. Jenkins, C. D., Zyzanski, S. J., Rosenman, R. H. 1976. Risk of new myocardial infarction in middle-aged men with manifest coronary heart disease. *Circulation* 53:342–47
34. National Heart, Lung, and Blood Institute, 1977. Forum on Coronary-Prone Behavior, St. Petersburg, Florida, June 1977. In press
35. Bartorelli, C., Zanchetti, A. eds. 1972. *Neural and Psychological Mechanisms in Cardiovascular Disease* pp. 1–379. Milan: Casa Editrice
36. Eliot, R. S. ed. 1974. *Stress and the Heart*, pp. 13–31, 69–141. Mt. Kisco, NY: Futura
37. Raab, W. ed. 1966. *Prevention of Ischemic Heart Disease*, pp. 31–168. Springfield, Ill.: Thomas
38. Raab, W., Chaplin, J. P., Bajusz, E. 1964. Myocardial necroses produced in domesticated rats and in well rats by sensory and emotional stresses. *Proc. Soc. Exp. Biol. Med.* 116:665
39. Rosenman, R. H., Friedman, M. 1974. Neurogenic factors in pathogenesis of coronary heart disease. *Med. Clin. North Am.* 58:269–79
40. Schneiderman, N. 1977 *Animal models relating behavioral stress and cardiovascular pathology: A position paper on mechanisms.* Presented at NHLBI Forum on Coronary-Prone Behavior. St. Petersburg, Florida, June 1977. DHEW-NHLBI. In press.
41. Myocardial infarct and other psychosomatic disturbances. 1968. *Psychother. Psychosom.* 16:189–292
42. Christian, P. 1968. Interdependenz von Umwelt und Person am Beispiel des Herzinfarktes, See Ref. 41, pp. 210–23
43. Lehman, E. W. 1967. Social class and coronary heart disease: a sociological assessment of the medical literature. *J. Chronic Dis.* 20:381–91
44. Caffrey, B. 1967. Factors involving interpersonal and psychological characteristics: a review of empirical findings. *Milbank Memorial Fund Q.* 45:119–39
45. Mai, F. M. M. 1968. Personality and stress in coronary heart disease. *J. Psychosom. Res.* 12:275–87
46. Mertens, C., Segers, M. J. 1971. L'influence des facteurs psychologiques dans la gènese des affections coronairennes. I. Données bibliographiques. *Bull Acad. R. Med. Belg.* 11:155–99
47. Syme, S. L., Borhani, N. O., Buechley, R. W. 1965. Cultural mobility and coronary heart disease in an urban area. *Am. J. Epidemiol.* 82:334–46
48. Bruhn, J. G., Wolf, S., Lynn, T. N., Bird, H. B., Chandler, B. 1968. Social aspects of coronary heart disease in a Pennsylvania German community. *Soc. Sci. Med.* 2:201–12
49. Shekelle, R. B., Ostfeld, A. M., Paul, O. 1969. Social status and incidence of coronary heart disease. *J. Chronic Dis.* 22:381–94
50. Syme, S. L., Hyman, M. M., Enterline, P. E. 1964. Some social and cultural factors associated with the occurrence of coronary heart disease. *J. Chronic Dis.* 17:277–89
51. Hinkle, L. E. Jr. Whitney, L. H., Lehman, E. W., Dunn, J., Benjamin, B., King, R., Plakun, A., Flehinger, B. 1968. Occupation, education, and coronary heart disease: risk is influenced more by education and background than by occupational experiences, in the Bell System. *Science* 161:238–46
52. Lehman, E. W., Schulman, J., Hinkle, L. E. Jr. 1967. Coronary deaths and organizational mobility: the 30-year experience of 1160 men. *Arch. Environ. Health* 15:455–61
53. Medalie, J. H., Kahn, H. A., Neufeld, H. N., Riss, E., Goldbourt, U., Perlstein, T., Oron, D. 1973. Myocardial infarction over a five-year period. I. prevalence, incidence and mortality experience. *J. Chronic Dis.* 26:63–84
54. Medalie, J. H., Snyder, M., Groen, J. J., Neufeld, H., Goldbourt, U.,

Riss, E. 1973. Angina pectoris among 10,000 men: 5-year incidence and univariate analysis. *Am. J. Med.* 55:583–94

55. Bruhn, J. G., Chandler, B., Miller, M. C. Wolf, S., Lynn, T. 1966. Social aspects of coronary heart disease in two adjacent, ethnically different communities. *Am. J. Public Health* 56:1493–1506
56. Bruhn, J. G., Chandler, B., Lynn, T. N., Wolf, S. 1966. Social characteristics of patients with coronary heart disease. *Am. J. Med. Sci.* 251: 629–37
57. Rahe, R. H., Lind, E. 1971. Psychosocial factors and sudden cardiac death: a pilot study. *J. Psychosom. Res.* 15:19–24
58. Theorell, T., Rahe, R. H. 1971. Psychosocial factors and myocardial infarction. I. An in-patient study in Sweden. *J. Psychosom. Res.* 15:25–31
59. Rahe, R. H., Bennett, L., Romo, M., Siltanen, P., Arthur, R. J. 1973. Subjects' recent life changes and coronary heart disease in Finland. *Am. J. Psychiatry,* 130:1222–26
60. Rahe, R. H., Paasikivi J. 1971. Psychosocial factors and myocardial infarction. II. An outpatient study in Sweden. *J. Psychosom. Res.* 15:33–39
61. Bengtsson, C., Hällström, T., Tibblin, G. 1973. Social factors, stress experience, and personality traits in women with ischaemic heart disease, compared to a population sample of women. *Acta Med. Scand.* (Suppl.) 549:82–92
62. Lundberg, U. Theorell, T., Lind, E. 1975. Life changes and myocardial infarction: individual differences in life change scaling. *J. Psychosom. Res.* 19:27–32
63. Theorell, T., Lind, E., Floderus, B. 1975. The relationship of disturbing life-changes and emotions to the early development of myocardial infarction and other serious illnesses. *Int. J. Epidemiol.* 4:281–93
64. Rahe, R. H., Arajarvi, H., Arajarvi, S., Punsar, S., Karvonen, M. J. 1976. Recent life changes and coronary heart disease in East versus West Finland. *J. Psychosom. Res.* 20:431–37
65. Weiss, N. S. 1973. Marital status and risk factors for coronary heart disease: the United States Health Examination Survey of Adults. *Br. J. Prev. Soc. Med.* 27:41–43
66. Bruhn, J. G., Paredes, A., Adsett, C. A., Wolf, S. 1974. Psychological predictors of sudden death in myocardial infarction. *J. Psychosom. Res.* 18:187–91
67. Williams, R. B. 1975. Physiological mechanisms underlying the association between psychosocial factors and coronary disease. In *Psychological Aspects of Myocardial Infarction and Coronary Care* ed, W. D. Gentry, R. B. Williams, pp. 37–50. St. Louis: Mosby
68. Keys, A., Taylor, H. L., Blackburn, H., Brozek, J., Anderson, J. T., Simonson, E. 1971. Mortality and coronary heart disease among men studied for 23 years. *Arch. Intern. Med.* 128:201–14
69. Thomas, C. B., Greenstreet, R. L. 1973. Psychobiological characteristics in youth as predictors of five disease states: suicide, mental illness, hypertension, coronary heart disease and tumor. *Johns Hopkins Med. J.* 132:16–43
70. Parkes, C. M., Benjamin, B., Fitzgerald, R. G. 1969. Broken heart: a statistical study of increased mortality among widowers. *Br. Med. J.* 1:740–43
71. Groen, J. J., Drory, S. 1967. Influence of psychosocial factors on coronary heart disease: a comparison of autopsy findings with the results of a sociological questionnaire. *Pathol. Microbiol.* 30:779–88
72. Henry, J. P., Ely, D. L., Stephens, P. M., Ratcliffe, H. L., Santisteban, G. A., Shapiro, A. P. 1971. The role of psychosocial factors in the development of arteriosclerosis in CBA mice. *Atherosclerosis* 14:203–18
73. Holmes, T. H., Rahe, R. H. 1967. The social readjustment rating scale. *J. Psychosom. Res.* 11:213–18

Ann. Rev. Med. 1978. 29:563–70
Copyright © 1978 by Annual Reviews Inc. All rights reserved

THE XYY GENOTYPE[1]

◆7298

Stanley Walzer, M.D.
Department of Psychiatry, University of Massachusetts Medical Center, Worcester, Massachusetts 01605

Park S. Gerald, M.D.
Departments of Pediatrics and Psychiatry, Childrens Hospital Medical Center, Boston, Massachusetts 02115

Saleem A. Shah, Ph.D.
Center for Studies of Crime and Delinquency, National Institute of Mental Health, Bethesda, Maryland 20014

INTRODUCTION

The XYY sex chromosome complement was first reported in 1961 by Sandberg et al (1) in a phenotypically normal male who was evaluated chromosomally (karyotyped) because he fathered a child with Down's syndrome. Between 1961 and 1965, the genotype was reported about 25 times in males with various physical abnormalities (most frequently hypogonadism), or in males karyotyped because they were related to patients with other chromosome abnormalities. In 1965 there was a meteoric rise of interest in the genotype following a report by Jacobs et al (2) that XYY males were identified with a frequency considered to be unexpectedly high (i.e. about 3% of the individuals studied) in a Scottish institution (Carstairs) for dangerous criminals, even though the actual incidence and prevalence of this genotype in the noninstitutionalized general background population were not yet known. Following this report there were numerous screenings of high-risk populations in mental, penal, and mental-penal (special security) settings, and the XYY karyotype—previously considered to be rare—was found to be one of the most common chromosome abnormalities in man (3). The extensive epidemiological literature that resulted from these ascertainment-biased screenings raised considerable speculation about a possible association between criminal behavior and the supernumerary Y chromosome. An XYY stereotype developed, which included tallness, low IQ, a behavior disability, and nodulocystic acne.

---

[1] Supported by USPHS Grants MH-17960 and HD-04807.

0066-4219/78/0401-0563$01.00

Since 1968 there have been increasing numbers of systematic investigations aimed at studying XYY males located through chromosome screenings of consecutively born male infants or of adult males in the general community. The intent of the studies has been to ascertain the background incidence and prevalence of the genotype and to understand the morphological, endocrinological, and behavioral variability that may be associated with it.

LIMITATIONS OF AVAILABLE DATA

The extensive literature on the XYY sex chromosome complement is considerably biased. Much of the information was obtained through screenings of populations selected *because* they demonstrated a particular personality or physical characteristic (i.e. antisocial behavior, tallness, etc) and individuals with this karyotype were found more frequently than expected. Such studies give no information about the developmental relevance of this genotype for chromosomally aberrant individuals not appearing in these specific populations.

Further bias resulted from the fact that most of the available data were obtained from individual case reports (often without suitable controls) and the behavioral descriptions of the affected males were anecdotal, impressionistic, and subjective.

A third source of bias stems from an unwarranted fear of investigations aimed at identifying biological contributants to behavioral development. There is concern that such studies will deemphasize the significance of complex psychosocial factors behind human behavioral development. Thus, authors often make emphatic and definitive statements that support or deny any relationship between the XYY genotype and behavioral development, even though the scientific evidence employed to defend their positions may be insufficient.

Systematic studies of XYY individuals ascertained through screenings of normal adults or of newborns will result in a less biased literature that will define more adequately the actual risk for an abnormal physical or psychological phenotype.

EPIDEMIOLOGY

The incidence of the XYY karyotype among the general newborn population (and the prevalence among the general adult population) is reported to be about 1 per 1000 males (4–6).

In extensive summarizations of the data published prior to 1972, Hook (3, 7) ascertained—in reports from 16 exclusively penal settings—that there were 26 XYY males found among the 5805 height-unselected males screened (1 per 225), or four times the background population frequency. In reports from 9 exclusively mental settings (with no special security requirements), Hook (7) ascertained 8 XYY males among the 2526 males screened (1 per 316), or about three times the background prevalence. In non-height-restricted studies of 3813 males in 20 mental-penal settings, Hook (3) ascertained 72 XYY males (1 per 53), or about 20 times the background frequency.

It has been suggested that the XYY genotype might occur more commonly among the progeny of lower socioeconomic groups, and that concomitant social and economic factors might account for their more frequent institutionalization. However, in newborn screenings for this abnormality in Edinburgh (8) and Boston (9), investigators found no significant difference in the social class distribution between the families of XYY infants and the total population from which they were drawn.

PHYSICAL CHARACTERISTICS

Most individual case reports emphasized tallness as being a rather typical phenotypic characteristic associated with the XYY sex chromosome complement. Further evidence for this association was provided by the significantly higher frequency of this genotype in institutionalized and "normal" males who are more than approximately 180 cm tall, as compared to similar samples of males less than 180 cm tall (5, 10, 11). When XYY males were compared with their blood relatives, they were frequently found to be taller than other members of their sibships and taller than their parents (5, 12, 13).

In systematic studies comparing XYY neonates with suitable controls, there have been no significant differences in the mean birth lengths of the two groups (14, and Walzer, unpublished data). Ratcliffe (14) reported that the XYY boys who became tall increased their stature at 2½–4 years of age.

It is possible that the physical characteristic of tallness makes XYY men with behavioral deviancy vulnerable to specific social decisions by the criminal justice system, and may lead to an increased probability of their being selected for confinement (15). However, Borgaonkar et al (13) demonstrated that the age-specific height distribution of institutionalized juvenile delinquents in Maryland did not differ significantly from the general background population. Therefore, tallness itself did not increase the likelihood of incarceration.

Nonspecific abnormalities in several body systems have been reported to be associated with the XYY genotype. Such diverse skeletal conditions as radioulnar synostosis, minor vertebral abnormalities, idiopathic osteoarthropathy, and Marfan's syndrome have been found occasionally in XYY males. The dermatologic characteristic most frequently reported is nodulocystic acne, which is found in 44% of known XYY males and in only 3% of XY males (16). A cardiac abnormality reported in association with the XYY karyotype is a prolonged PR interval (17, 18); however, other investigators have been unable to confirm this finding (12, 19, 20).

Although a variety of neurological conditions have been reported in individual instances of this genotype, the most commonly reported variations are in the electroencephalographic (EEG) recordings. Borgaonkar & Shah (12) summarized the case reports published prior to 1971 and found 28 reports of XYY males with normal EEGs and 19 with abnormal recordings. The most frequently reported abnormality in both institutionalized and "normal" XYY men is a generalized slowing of the background electrical activity (21, 22). Since slowness of the EEG frequency is normally prominent in young age groups, it has been suggested that the supernumerary Y chromosome might retard brain maturation (22). Delayed speech development and/or problems with articulation have been reported in XYY males (14, 23, 24).

Although abnormalities in the genitourinary system (i.e. hypogonadism) were reported frequently in the early literature on the XYY genotype, the majority of later-identified XYY males have been found to have normal-appearing external male development. However, abnormalities of testicular histology have been reported frequently, including abnormal seminiferous epithelium, and impairment of spermatogenesis (11, 25). In a few instances, normal testicular histology has been reported (25, 26).

Investigations into the endocrinologic status of XYY individuals usually revealed the plasma testosterone levels to be within the normal range or comparable to those obtained from suitably matched controls (27). Although the plasma pituitary luteinizing-hormone levels were frequently elevated, the other plasma pituitary gonadotrophin levels were usually normal. Papanicolaou (28) conjectured that in XYY males the Leydig cells of the testes are relatively insensitive to stimulation by luteinizing hormone, and therefore the pituitary produces excessive quantities of this hormone in order to overcome the refractoriness.

PSYCHOLOGICAL CHARACTERISTICS

Intelligence

There is considerable speculation about the intelligence levels of XYY individuals. Although the general distribution of IQ scores in institutionalized XYY males is lower than that expected, a wide range of individual IQ scores has been reported—from superior to low. Since institutionalized populations are biased toward lower IQ scores, no conclusions about the relative intelligence of XYY men can be drawn from these data.

Studies of institutionalized XYY males in which matched XY controls were drawn from the same institution resulted in variable findings. Griffiths et al (29) and Griffiths (30) found significantly higher intelligence levels in the normal controls, while other investigators found no significant difference in mean IQ between XYY individuals and controls (31, 32).

Nielsen & Christensen (11) reported that the Full-Scale Wechsler IQ scores of 9 noninstitutionalized (as well as 17 institutionalized) XYY males were significantly lower than expected (according to the 1958 Wechsler IQ norms). Witkin et al (33), having ascertained 12 XYY males from a cohort of 4139 height-selected "normal" Danish males, systematically collected data relating to convictions for previous criminal offenses, levels of intellectual functioning (as indicated by scores on an Army selection test and by educational achievement), and the parental social class at the time of the subject's birth for almost all men in the study. These investigators found a lower overall intelligence level for the XYY group compared to XY controls.

The IQ scores reported for XYY children followed prospectively from birth are too few in number to allow any conclusions at this time. However, Ratcliffe (14) reported IQ scores ranging from 78–130+ (with a mean IQ score of 99.5) on 14 XYY infants and children; Valentine (34) reported normal intelligence levels (IQs of 102–106) for three XYY infants and a borderline level (IQ of 84) for one. Grant & Hamerton (35) reported normal development assessments (Yale Developmental Schedules) on four XYY infants.

Psychosocial Development

Individual instances of the XYY karyotype have been reported in association with a wide variety of psychiatric diagnoses and behavioral descriptions (36). Since the vast majority of XYY individuals reported in the literature were located through chromosome screenings of behaviorally abnormal populations, it is to be expected that these chromosomally abnormal men frequently manifest aberrant social or behavioral development.

The epidemiological data showing a significant overrepresentation of XYY males in mental-penal settings raised questions about the frequency and nature of criminal behavior in individuals with a supernumerary Y chromosome. Studies comparing crime rates prior to institutionalization for XYY males and suitable controls have been inconclusive (38, 39). Similarly, in studies of XYY males ascertained in the general population, Noël et al (5) found no increase in crime rate for the XYY males as compared to XY controls, while Witkin et al (33) found a significantly increased crime rate for the XYY men.

Although several investigators reported a significantly younger mean age of first conviction in institutionalized XYY males compared to suitable controls drawn from the same populations (40), other investigators were unable to find such a difference (29, 37, 38, 42).

Several reports emphasized that the crimes committed by XYY males were not related to "aggression" (i.e. nonsexual violence against persons) but were usually sexual offenses and crimes against property (12, 29, 32, 33, 37, 41, 42).

Witkin et al (33) suggested an intellectual dysfunction hypothesis to explain the elevated crime rate they found in 12 XYY males ascertained in a cohort of height-selected normal Danish males. The XYY subjects had a lowered intellectual capability and a higher level of criminality than XY controls. Since the intelligence scores of controls were shown to correlate with the frequency of occurrence of registered crimes leading to convictions, the investigators suggested that the lowered intelligence capacity of XYY males was an important mediating variable to explain the higher mean rate of criminal convictions. However, the correlation between low intelligence and criminality was insufficient to explain fully the elevation in XYY crime rate; further studies are required to test the intellectual-dysfunction hypothesis.

Although the epidemiological studies suggest that there is a significantly increased risk for XYY males to appear in mental-penal settings, the data give no information about the developmental relevance or irrelevance of this genotype for the vast majority of XYY individuals not appearing in these social settings. Unbiased information is only now becoming available through systematic studies of XYY individuals located among the general adult male population, or through long-term longitudinal studies of XYY males ascertained in newborn screenings and followed prospectively throughout their developmental years.

Employing a systematic double-blind psychological assessment, Noël et al (5) were able to distinguish between matched controls and XYY males drawn from the same general male population. Eight of the 14 XYY men showed a predominance of immediate gratification test responses together with impulsivity when faced with emotional stimuli. There was also a lack of emotional control, with frequent histories of emotional outbursts (i.e. "fits of temper") when faced with frustration. Six of the 14 XYY subjects were hypercontrolled and rigidly defended against aggressive impulses. The XYY men with below average IQ scores and unstable personalities appeared to have a lower threshold for the control of aggression in frustrating or provocative situations. Thus, XYY individuals living apparently normal lives in the community could be distinguished from non-XYY men on the basis of formal psychological assessment.

As a group, XYY males may be at increased risk for some behavioral disability; however, this may not appear in a specific XYY individual. Furthermore, if it does develop, the factors responsible for its occurrence are not known at this time.

Longitudinal prospective studies of XYY infants and children are now underway at several centers around the world, and will contribute information about the developmental variability encountered in young children with this genotype. It is possible that certain intrinsic factors related to the presence of a supernumerary Y chromosome may limit somewhat the child's adaptive capacities, and contribute to the development of behavior or a temperamental style that puts stress on the parental coping strategies and makes the child difficult to manage. However, it is probable that the dynamics of family life and the child-rearing practices (related to such factors as the parents' educational levels, social class and ethnic backgrounds) determine ultimately whether the behavioral problems of infancy or early childhood expand into major handicapping difficulties or become self-regulated, allowing the affected individual an opportunity for developmental competence (24).

There is evidence that the social management of the behavioral deviancy seen in some XYY individuals may be class related; XYY men from the upper class and upper-middle class avoid institutionalization while those individuals from the lower-middle class and lower class do not (24).

Since it is probable that there is considerable variability in the phenotypic development of XYY males, it is inappropriate to allude to an XYY syndrome. The term syndrome implies a degree of symptom consistency that is not supported by the data available at this time (12). Furthermore, since we lack an adequate system for classifying human aggressive behaviors and determining the base rates for different aggressive behaviors in the background population, we cannot state definitely whether there is an increase in aggressive behavior associated with this genotype (27).

CONCLUSION

The epidemiological data pertaining to the XYY genotype suggest that there is a three- to fourfold overrepresentation of XYY individuals in mental and penal settings and a twentyfold overrepresentation in mental-penal (special security) settings. The reasons behind the risk for behavioral disability are not known at this time. Tallness and nodulocystic acne are reported frequently to be associated with the genotype. Since much of the available information about the XYY sex chromosome complement is biased, more data is required before definitive statements can be made about the personality characteristics or intelligence of the vast majority of XYY men not appearing in social settings oriented towards behavioral deviancy.

Literature Cited

1. Sandberg, A. A., Koepf, G. F., Ishihara, T., Hauschka, T. S. 1971. An XYY human male. *Lancet* 2: 488–89
2. Jacobs, P. A., Brunton, M., Melville, M. M., Brittain, R. P., McClemont, W. F. 1965. Aggressive behavior, mental subnormality and the XYY male. *Nature* 208: 1351
3. Hook, E. B. 1973. Behavioral implications of the human XYY genotype. *Science* 179:139–50
4. Jacobs, P. A., Melville, M. M., Ratcliffe, S. G., Keay, A. J., Syme, J. 1974. A cytogenetic survey of 11,680 newborn infants. *Ann. Hum. Genet.* 37:359–76
5. Noël, B., Duport, J. P., Revil, D., Dusseyer, I., Quack, B. 1974. The XYY syndrome: Reality or myth? *Clin. Genet.* 5:387–94
6. Walzer, S., Gerald, P. S. 1977. In *Population Cytogenetics,* ed E. B. Hook, I. H. Porter, pp. 45–61. New York: Academic. 374 pp.
7. Hook, E. B. 1975. Rates of XYY genotype in penal and mental settings. *Lancet* 1:98
8. Ratcliffe, S. G., Evans, H. J. 1974. Sex-chromosome abnormalities and social class. *Lancet* 1:1144
9. Walzer, S., Gerald, P. S. 1975. Social class and frequency of XYY and XXY. *Science* 190:1228–29
10. Baker, D. 1972. Chromosome errors and antisocial behavior. *Crit. Rev. Clin. Lab. Sci.* 3:41–101
11. Nielsen, J., Christensen, A. L. 1974. Thirty-five males with double Y chromosome. *Psychol. Med.* 4: 28–37
12. Borgaonkar, D. S., Shah, S. A. 1974. The XYY chromosome: Male or syndrome? *Prog. Med. Genet.* 10:135–222
13. Borgaonkar, D. S., Unger, W. M., Moore, S. M., Crofton, T. A. 1972. 47,XYY syndrome, height and institutionalization of juvenile delinquents. *Br. J. Psychiatry* 120:549–50
14. Ratcliffe, S. G. 1976. The development of children with sex chromosome abnormalities. *Proc. R. Soc. Med.* 69:189–191
15. Shah, S. A., Roth, L. H. 1974. In *Handbook of Crimonology,* ed. D. Glaser, pp. 101–73. Chicago:Rand McNally. 1180 pp.
16. Voorhees, J. J., Wilkins, J. W., Hayes, E., Harrell, E. R. 1972. Nodulocystic acne as a phenotypic feature of the XYY genotype. *Arch. Dermatol.* 105:913–19
17. Price, W. H., Wilson, J. 1971. Electrocardiograph studies in individuals with different sex chromosome complement. *Excerpta Med. Int. Cong. Hum. Genet., 4th.* 233:531 (Abstr.)
18. Vianna, A. M., Frota-Pessoa, O., Lion, M. F., Decourt, L. 1972. Searching for XYY males through electrocardiograms. *J. Med. Genet.* 9:165–67
19. Char, F., Borgaonkar, D. S. 1971. Electrocardiogram in males with the 47, XYY karyotype. *Lancet* 1:1242
20. Steiness, E., Nielsen, J. 1970. The electrocardiogram in males with the 47, XYY karyotype. *Lancet* 1:1402–3
21. Fenton, G. W., Tennent, T. G., Comish, K. A., Rattray, N. 1971. The EEG and sex chromosome ab-

normalities. *Br. J. Psychiatry* 119: 185–90

22. Volavka, J., Mednick, S. A., Sergeant, J., Rasmussen, L. 1977. Electroencephalograms of XYY and XXY men. *Br. J. Psychiatry* 130: 43–47
23. Leonard, M. F., Landy, G., Ruddle, F. H., Lubs, H. A. 1974. Early development of children with abnormalities of the sex chromosomes: A prospective study. *Pediatrics* 54:208–12
24. Money, J., Franzke, A., Borgaonkar, D. S. 1975. XYY syndrome, stigmatization, social class and aggression: Study of 15 cases. *South. Med. J.* 68:1536–42
25. Skakkebaek, N. E., Hulten, M., Jacobsen, P., Mikkelsen, M. 1973. Quantification of human seminiferous epithelium. II. Histological studies in eight 47, XYY men. *J. Reprod. Fertil.* 32:391–402
26. Shapiro, L. R. 1970. Hormones and the XYY male. *Lancet* 1:1347
27. Meyer-Bahlburg, H. F. L. 1974. In *Sex Differences in Behavior*. ed. R. C. Friedman, R. M. Richart, R. L. VandeWiele, pp. 433–53. New York: Wiley. 495 pp.
28. Papanicolaou, A. D., Kirkham, K. E., Loraine, J. A. 1968. Abnormalities in urinary gonadotrophin excretion in men with a 47, XYY sex chromosome constitution. *Lancet* 2:608–10
29. Griffiths, A. W., Richards, B. W., Zaremba, J., Abramowicz, T., Stewart, A. 1970. Psychological and sociological investigation of XYY prisoners. *Nature* 227:290–92
30. Griffiths, A. W. 1971. Prisoners of XYY constitution:Psychological aspects. *Br. J. Psychiatry*. 119:193–94
31. Hope, K., Phillips, A. E., Loughran, J. M. 1967. Psychological characteristics associated with XYY sex chromosome complement in a state mental hospital. *Br. J. Psychiatry*. 113:495–98
32. McKerracher, D. W. 1971. Psychological aspects of a sex chromatin abnormality. *Can. Psychol.* 12:270–81
33. Witkin, H. A., Mednick, S. A., Schulsinger, F., Bakkestrøm, E., Christiansen, K. O., Goodenough, D. R., Hirschhorn, K., Lundsteen, C., Owen, D. R., Philp, J., Rubin, D. B., Stocking, M. 1976. Criminality in XYY and XXY men. *Science* 193:547–55
34. Valentine, G. H., McClelland, M. A., Sergovich, F. R. 1971. The growth and development of four XYY infants. *Pediatrics*. 48:583–84
35. Grant, W. W., Hamerton, J. L. 1976. A cytogenetic survey of 14,069 newborn infants II. Preliminary clinical findings on children with sex chromosome anomalies. *Clin. Genet.* 10:285–302
36. Owen, D. R., 1972. The 47,XYY male: a review. *Psychol. Bull.* 78: 209–33
37. Casey, M. D., Blank, C. E., McLean, T. M., Kohn, P., Street, D. R. K., McDougall, J. M., Gooder, J., Platts, J. 1973. Male patients with chromosome abnormality in two state hospitals. *J. Ment. Defic. Res.* 16:215–16
38. Griffiths, A. W., Richard, B. W., Zaremba, J., Abramowicz, T., Stewart, A. 1969. In *Criminological Implications of Chromosome Abnormalities*, ed. D. J. West, pp. 32–43. Cropwood Round-Table Conference, Inst. Criminol. Univ. Cambridge, England
39. Clark, G. R., Telfer, M. A., Baker, D., Rosen, M. 1970. Sex chromosome, crime, and psychoses. *Am. J. Psychiatry*. 126:1659–63
40. Nielsen, J. 1970. Criminality among patients with Klinefelter's syndrome. *Br. J. Psychiatry*. 117:365–69
41. Price, W. H., Whatmore, P. B. 1967. Criminal behavior and the XYY male. *Nature* 213:815
42. Street, D. R. K., Watson, R. A. 1969. See Ref. 38, pp. 61–67

Ann. Rev. Med. 1978. 29:571–82
Copyright © 1978 by Annual Reviews Inc. All rights reserved

CARDIAC ASSIST DEVICES ◆7299

John J. Lamberti, M.D. and Leon Resnekov, M.D.[1]
Departments of Surgery and Medicine, The University of Chicago Pritzker School of Medicine, Chicago, Illinois 60637

INTRODUCTION

During the last twenty years, substantial progress has occurred in the treatment of cardiac failure due to valvular heart disease or following acute myocardial infarction with or without left ventricular aneurysm. Cardiopulmonary bypass with conventional bubble oxygenators has been developed to the point where two hours of bypass rarely result in significant mortality or morbidity. Despite these advances there remains a large group of patients in whom left ventricular function has been irreversibly damaged as a consequence of long-standing untreated valvular heart disease. In addition, patients with severe chronic ventricular failure as a result of advanced coronary heart disease or congestive cardiomyopathy are not usually candidates for corrective heart surgery. It has been estimated that more than 400,000 patients die each year as a result of ineffective cardiac pumping action (1). Although there have been dramatic successes following cardiac transplantation, mechanical assist devices capable of supporting left ventricular function for prolonged periods of time are still in the developmental and experimental stages.

During the next ten years, a variety of devices to prolong life in patients with end-stage myocardial failure will become available. Successful short-term circulatory assistance in man has already been achieved with several devices. All of these techniques, however, are relatively invasive and therefore prone to complications.

Left ventricular assist devices may be classified as in Table 1. In general, cardiac assistance is achieved using mechanical devices to replace the malfunctioning left ventricle or even the entire heart. The alternate approach is orthotopic or heterotopic cardiac allografting. Experience with cardiac transplantation is limited; encouraging reports, however, suggest that long-term palliation can often be achieved with this technique.

---

[1] Supported in part by the USPHS Grant IP17 HL 17648, Special Center of Research in Ischemic Heart Disease (SCOR–IHD), USPHS Contract NO1-HL-81334, Myocaridal Infarction Research Unit (MIRU), and by the Chicago Heart Association.

0066-4219/78/0401-0571$01.00

Table 1 Cardiac assist devices

Mechanical	Transplantation
Cardiopulmonary Bypass	Orthotopic
Conventional oxygenator	Heterotopic (parallel)
Membrane oxygenator	
Left ventricular bypass	
Counterpulsation	
Intra-aortic balloon counterpulsation	
Dynamic aortic patch	
External pressure counterpulsation	
Artificial Heart	
Total	
Left ventricular assist	

In this report we review currently available techniques and discuss their merits, problems, and possibilities for the future.

CARDIAC TRANSPLANTATION

Following the pioneering work of Carrel & Guthrie (2), many investigators explored the possibility of cardiac transplantation in the experimental laboratory. In 1959 Lower & Shumway (3) reported the first successful orthotopic cardiac transplantation in the dog. Subsequent work has been devoted to developing methods for detecting and preventing allograft rejection. In 1967, the first human heart transplant generated world-wide interest and enthusiasm in the technique (4). Within a year, 101 transplant operations had been performed (5). Thereafter, a gradual decline in enthusiasm for this technique occurred. Successful homotransplantation of the heart requires more than a willing recipient, a suitable donor, and a skilled operating team. Technical details of the operation were well established by Lower & Shumway (3) and a number of centers achieved early clinical success with transplantation. Nevertheless the problem of detecting and treating transplant rejection quickly turned early success into disappointment and failure. The group at Stanford University continued to develop the technique of cardiac transplantation and now reports excellent clinical success (6).

Technical aspects of transplantation have varied little in the past six years (5). The donor heart is excised and cooled to approximately 10°C by immersing in cold saline. Removal of both donor and recipient hearts at the midatrial level facilitates anastomosis of the atria. After completion of anastomosis of the great vessels, and recovery of the graft, cardiopulmonary bypass is gradually discontinued. Hemodynamic studies in humans during the early postoperative period show that the pumping performance of the orthotopically transplanted (denervated) heart is significantly depressed for three to four days. Isoproterenol in standard doses can normalize graft function and is used routinely in all recipients for two to four days. In addition, pacing of the donor heart via temporary wires is occasionally necessary. Maintenance immunosuppressive regimen consists of corticosteroids, azathioprine, or cyclophosphamide, and antithymocyte globulin. In general, episodes of rejection can be expected within the first eight postoperative weeks. Rejection is thought to be mediated predominantly by cellular mechanisms involving thymus-derived lymphocytes. The rejecting heart shows edema, interstitial infiltration with mononuclear cells, hemorrhage, endothelial and endocardial damage, and occasionally acute arteritis. These pathological changes may be expressed electrophysiologically as a generalized decrease in electrocardiographic voltage, artrial dysrhythmias, and S-T and T wave changes on the electrocardiogram, or they may be expressed hemodynamically, as impaired left ventricular function. Overt failure, however, is a relatively late manifestation of graft rejection and is usually associated with some degree of irreversible myocardial damage. The new technique of serial endocardial biopsy of the cardiac graft permits direct histological diagnosis of rejection and allows assessment of the response to immunosuppressive treatment; it has proved highly valuable in managing recipients after transplantation (7). Unfortunately, the immunosuppressive regimen necessary to prevent rejection decreases host resistance to infection, which has caused or contributed to two thirds of all postoperative deaths in the Stanford series. Maximal survival after cardiac transplantation now exceeds five years, but only in a few cases. In the total Stanford series, actuarially calculated survival to six months, one year, and two years postoperatively stands currently at 51%, 44%, and 38%. More recently, reported survival rates have been somewhat higher (6). It is also notable that one-year survival in recipients who have previously undergone open heart operations for coronary artery or rheumatic disease is approximately 80%. The reason for this relatively favorable status enjoyed by previously operated patients has not yet been defined. It has been suggested, however, that it may be related to immunologic factors such as protective antibodies developed in response to multiple blood transfusions.

Long-term problems faced by cardiac transplant recipients relate to the hazards of continued-maintenance immunosuppression; these include infection, lipid metabolism disorders, and osteoporosis. In addition, a form of accelerated atherosclerosis is also known to occur in the donor coronary arteries. The incidence of these coronary atherosclerotic changes has declined markedly as a result of long-term treatment with oral anticoagulation and platelet antagonists, and control of abnormal serum lipid levels (6).

In response to some of the problems associated with orthotopic cardiac allografting, Barnard & Losman (8) of the University of Cape Town developed a technique for bypassing the donor's diseased left ventricle with a cardiac allograft in parallel with the recipient heart. An apparent advantage of this technique of "piggy-back" cardiac transplantation is that the recipient is not dependent solely on the allograft for cardiac output. During treatment of rejection, the patient may retain a basal cardiac status with function of his own severely diseased heart. In addition, severe pulmonary hypertension is not a definite contraindication to left ventricular assistance by parallel cardiac transplantation. A comparison of electrocardiographic voltage from the donor and the recipient hearts may yield information regarding early signs of allograft rejection. On the other hand, endomyocardial biopsy is not yet possible in these heterotopic allografts. Therefore, early rejection may be more difficult to detect, although these patients seem to tolerate rejection episodes better than patients totally dependent on their allograft for myocardial function. The Cape Town series is still too small to draw long-term conclusions. It would seem that for certain recipients, "double heart" transplantation offers a useful mechanism of palliation.

A variety of factors make orthotopic cardiac transplantation an unlikely mechanism for salvaging large numbers of patients dying of left ventricular failure. Potential cardiac recipients must satisfy certain basic criteria (5). They must have advanced progressive cardiac disease intractable to ordinary forms of treatment and associated with an estimated survival of weeks to months. They should be free of active infectious disease or overt diabetes mellitus. Finally, they must not have severely elevated pulmonary vascular resistance as a result of chronic left ventricular failure. Severe pulmonary hypertension in the recipient may not be overcome by implanting a normal donor right ventricle. In addition, it is preferred that recipients be under the age of 55 years.

Criteria for the donor are similar to those applied in the case of other organ transplants. The donor should be less than 40 years of age and free of known cardiac disease. The presence of malignancy or infection precludes transplantation. In general, victims of head trauma with irreversible brain damage are ideal candidates for cardiac donors.

In summary, cardiac transplantation, either orthotopic or heterotopic, offers a limited form of long-term palliation for a small number of patients with end-stage left ventricular failure. Even if the problem of allograft rejection is solved, the supply of donor organs appears limited in comparison to the clinical need.

MECHANICAL LEFT VENTRICULAR ASSISTANCE

Cardiopulmonary Bypass

Conventional cardiopulmonary bypass, employing roller pumps and any of a variety of types of bubble oxygenators, can be used safely for two to three hours. Continued use therafter results in trauma to formed elements in the blood with resultant hemolysis and thrombocytopenia. The conventional type of oxygenator first used in 1953 by Gibbon (9) is therefore only a very temporary means of support for the failing ventricle. In patients dying of aortic stenosis or with acute coronary occlusion occurring in the cardiac catheterization laboratory, prompt initiation of cardiopulmonary bypass and corrective surgery may be all that is necessary to restore adequate left ventricular function.

Prolonged bypass with roller pumps and membrane oxygenators is successful in the treatment of pulmonary insufficiency due to a variety of causes (10). On several occasions, prolonged extracorporeal membrane oxygenation has been used to support patients with primary myocardial failure (11). Although this technique is cumbersome and has high morbidity, predominantly bleeding, it may be a useful form of short-term assistance in selected cases. Patients undergoing open heart surgery with a membrane oxygenator are those most likely to benefit from this form of relatively prolonged ventricular assistance following operation. There remains disagreement over the amount of bypass needed to significantly reduce myocardial oxygen requirements.

Salisbury et al (12) found that approximately 80–90% of the cardiac output had to be bypassed in a venoarterial assistance system before oxygen consumption of the left ventricle was reduced. This is hardly surprising, since the oxygen consumption of the left ventricle is directly related to the pressure generated during systole but not to the cardiac output (13). Recently, Litwack (14) suggested that prolonged left ventricular assistance could be obtained by withdrawing blood from the left atrium and returning it to the aorta with a roller pump. The advantage of this system is left ventricular bypass without an oxygenator in circuit, as well as the return of oxygenated blood to the proximal aorta. In the experience of Litwack, a large volume flow following completion of the surgical procedure is essential for short-term survival in patients suffering refractory left ventricular failure. Cannulae that may be occluded by a mandrill or obturator are connected to the left atrium and the aorta and then buried under the skin. Implantation requires a thoracotomy, but discontinuation may be performed under local anesthesia.

Transapical left ventricular bypass is a variation of this technique (15). Following the completion of operation in patients suffering refractory cardiac failure, a cannula is inserted through the apex of the left ventricle. Blood is removed from the left ventricle continuously and returned to the aorta with a roller pump. As in the Litwack procedure, the apparatus necessary for this type of bypass is simple and available in all cardiac surgical operating rooms. In both cases the chest may be closed and the patient taken to the Intensive Care Unit while on partial bypass. Because there is no oxygenator in the system, trauma to the blood elements is minimal if the roller pump is properly calibrated. These techniques are useful in those few patients who remain in severe left ventricular failure following completion of cardiac surgery.

Counterpulsation

The concept of counterpulsation was originally proposed by Kantrowitz in 1953 (16) and initially employed by Clauss et al in 1961 (17). In theory, withdrawing blood from the arterial system during systole and reintroducing it into the arterial system during diastole should increase the cardiac output while decreasing the external work of the heart. The reduction in systolic pressure decreases left ventricular afterload, thereby decreasing left ventricular wall tension and consequently myocardial oxygen consumption. Elevation of diastolic blood pressure increases coronary perfusion pressure and presumably increases coronary blood flow. Clauss et al used a reciprocating pump triggered by the electrocardiogram to withdraw and reinject blood into the femoral artery. Trauma to formed elements in the blood was high with the early devices, which required cannulation of the femoral artery and vein. Moulopoulos, Topaz & Kolff (18) simplified the system considerably by introducing intraaortic balloon counterpulsation (IABCP). This method of circulatory assistance uses a balloon catheter, which is introduced through the femoral artery and passed retrograde into the descending thoracic aorta. With precise synchronization to the patient's electrocardiogram, inflation of the balloon during early diastole results in an increased diastolic arterial pressure. Prompt deflation of the balloon before ventricular systole results in a fall in intraaortic pressure. The left ventricle ejects, therefore, against a reduced aortic pressure, which decreases impedance and afterload. The compromised left ventricle is therefore able to eject a larger volume of blood against a decreased pressure load, resulting in less work. Left ventricular end-diastolic pressure and volume are lessened. Numerous experimental and clinical reports document that diastolic augmentation effectively improves circulatory function (19, 20).

IABCP is currently the most widely used form of left ventricular assistance. Patients with left ventricular power failure after acute myocardial infarction are at high risk of sudden death, and cardiogenic shock treated with drug support alone carries a mortality of 90%. The indication for the use of counterpulsation is acute myocardial depression caused by a potentially reversible condition (Table 2).

In our own series at the University of Chicago, more than 100 patients have been treated with IABCP since September 1973. Survival has depended on the degree of irreversible damage to the left ventricle in the individual patient. In the last year, ten patients with cardiogenic shock or refractory left ventricular failure underwent insertion of IABCP prior to cardiac catheterization. Five (50%) survived corrective surgery and were discharged from the hospital. This group of patients is generally considered to have a 90–95% mortality without counterpulsation and corrective surgery. Similar results have been reported by other groups (21, 22).

Patients suffering refractory left ventricular failure at the completion of cardiopulmonary bypass also benefited from IABCP. With IABCP, 60% of patients who could not be successfully weaned from cardiopulmonary bypass

Table 2 Indications for intraaortic balloon counterpulsation

1. Left ventricular power failure before operation (potentially correctable cardiac lesions)
2. Left ventricular power failure after cardiopulmonary bypass
3. Refractory dysrhythmias after cardiopulmonary bypass (usually associated with left ventricular power failure)
4. Prophylactic insertion before surgery (patients with severe left ventricular dysfunction or severe unstable angina)
5. Other considerations:
    A. Pulsatile cardiopulmonary bypass
    B. To improve flow in bypass grafts in the immediate postoperative period.

despite full pharmacologic assistance were nevertheless weaned from bypass and experienced long-term survival. Patients with unstable angina pectoris refractory to medical management may be stabilized with IABCP and then undergo complete cardiac catheterization and coronary angiography in safety. Surgery may then be performed if sensible, while the intraaortic balloon is still in position. We had 100% survival in patients with unstable angina managed in this fashion.

A final group of patients in whom IABCP is useful are those with severe chronic depression of left ventricular function due to coronary artery disease. Previous reports suggested that the operative mortality after coronary artery bypass grafting is high in those with severe depression of left ventricular ejection fraction (23). We found that operative mortality for such high-risk patients (ejection fraction ≤ 30%) may be reduced to between 10 and 20% with the use of prophylactic IABCP before the induction of anesthesia, during operation, and in the early postoperative period.

IABCP is an invasive method of left ventricular assistance. An incision in the groin is used to insert the balloon cannula through the femoral artery, and flow to the distal artery has to be preserved. Balloon counterpulsation may be used for seven to ten days with minimal complications in most cases. On occasion we used IABCP continuously for more than two weeks.

In summary, IABCP may be considered as a form of temporary left ventricular assistance that is useful in a large number of patients suffering potentially reversible left ventricular dysfunction. However, this technique is not useful for the long-term treatment of patients with irreversible damage to the left ventricle and inadequate left ventricular function.

External Pressure Circulatory Assist

External pressure circulatory assist (EPCA) was introduced by Dennis (24). In this form of diastolic augmentation, the patient's legs are enclosed in a rigid plastic box and surrounded by a water-filled plastic bag. The patient's electrocardiogram triggers a hydraulic system that inflates and deflates the bag, which compresses the legs during diastole. This results in an increase in aortic diastolic pressure, thereby increasing coronary artery perfusion pressure. Compression of the leg veins increases venous return to the heart, which is probably why EPCA increases the cardiac output.

We evaluated the hemodynamic effects of 60 minutes of EPCA in 15 patients with documented acute myocardial infarction, low cardiac index, and left ventricular failure (20). Studies were performed within 24 hours of admission to our Coronary Care Unit. Evaluation of the effects of EPCA revealed that diastolic augmentation can be consistently produced using the device. There is an increase in cardiac output, particularly in patients in left ventricular failure. No beneficial change, however, in left ventricular preload (pulmonary artery wedge pressure), or afterload (aortic systolic pressure) occurred; thus any improvement in cardiac output resulted from an increased venous return, and unlike IABCP there was no reduction in external work of the heart.

EPCA is thus a noninvasive means of improving cardiac output in patients with acute left ventricular decompensation. The apparatus is cumbersome and is difficult to use during cardiac catheterization and while performing coronary artery bypass surgery.

In summary, EPCA is a useful technique for improving the hemodynamic status of patients with severe left ventricular dysfunction. It is not, however, suitable for long-term left ventricular assistance, nor is it flexible enough to use in the management of cardiogenic shock patients who are undergoing cardiac catheterization and corrective cardiac surgery.

Dynamic Aortic Patch

The dynamic aortic patch (DAP) was developed by Kantrowitz (25) as an extension of the concepts validated by IABCP. A balloon-like device is sutured into the wall of the descending aorta under cardiopulmonary bypass. The hemodynamic actions of this prosthesis are essentially the same as those of IABCP. A console similar to that used for IABCP drives the patch in synchrony with the electrocardiogram. The intra- and extravascular surface of the pumping chamber and tube are covered with Dacron® velour, which permits tissue ingrowth and is thought to enhance the growth of a stable pseudointimal layer in the blood stream, thus lessening the chance of embolism. The power source and electronics for this device are not implantable; accordingly, a Dacron velour tube is led out through the skin. The device works well in the experimental laboratory and was successfully implanted in a patient. One patient was discharged from the hospital with a functioning DAP. He died three months after implantation with an infection around the pumping chamber. The dynamic aortic patch was intact and its intravascular surface had developed a glistening continuous layer of fibrin. There were no signs of embolization. This technique is known to be hemodynamically effective and might represent a method for long-term ventricular assistance. A fully implantable power source and control mechanism must be developed for it to be used as a long-term ventricular assist device.

ARTIFICIAL HEART

The development of a total artificial heart (TAH) has been studied since 1937, when Demikhov devised a compact apparatus consisting of two adjacent pumps that performed the function of both ventricles of the dog heart as given by Norman (26). Since then, a variety of devices have been developed to replace or assist the failing left ventricle. Total heart replacement requires removal of the diseased organ and renders the recipient dependent on his new heart for survival. In contrast, partial replacement or left ventricular assistance with a device that supplements or replaces the left ventricle has been proposed as an alternative to total heart replacement. With a left ventricular assist device removal may be possible, if left ventricular function recovers or if the residual myocardium adapts to its chronic work load. Both types of assist devices face similar developmental problems. The abdominal left ventricular assist device appears closer to clinical use, since it does not replace the recipient's heart and therefore can be removed after prolonged pumping and following clinical recovery.

Several power sources have been considered for artificial hearts. Electric motors, electromagnets, and pneumatic engines have all been advocated and tried (27). Nuclear-powered artificial hearts are currently under development in a number of centers (28), but heat dissipation remains a major problem. Kolff (28) suggested that a patient requiring TAH should first have an electromotor blood pump implanted, and later convert to a nuclear driven unit if long-term survival appears possible.

Most currently available power sources require percutaneous tubing and wiring with the attendant risk of infection. Schuder (29) suggested that electric energy sufficient to drive an artificial heart could be transmitted through the skin without perforating it. This technique is presently under evaluation. Total artificial hearts activated by transcutaneous external pneumatic-drive lines successfully kept calves alive for up to three months (28), although the stroke volume and cardiac output were relatively low. In 1969, Cooley, Liotta et al (30) implanted an externally activated, pneumatically driven, total mechanical heart into the mediastinum of a patient. The device sustained this man for approximately 64 hours, then a biologic heart transplant was performed. The patient died 36 hours later. Since that time, most activity in the development of a TAH has been confined to the research laboratory.

Control of the TAH is a subject of much interest. Kolff believes that for basal activity levels, Starling's law of the heart will operate (28). In some of the calves with total artifical hearts studied at the University of Utah, a high venous oxygen saturation was noted in the mixed venous blood. This suggests that the resting output of the TAH is probably greater than necessary. These animals were able to exercise daily without evidence of distress. A variety of servomechanisms have been proposed for regulating the cardiac output of the TAH. Currently, it appears that cardiac output regulation is not the major factor that precludes the clinical application of the technique.

The size and shape of a TAH are major considerations in development. Adaptation of the venous connections of the heart to the natural structures must ensure free movement of the subject without kinking or obstruction to venous return. The Jarvik heart (31) has a very small ventrodorsal diameter and functions successfully in the laboratory without producing signs of right heart failure. In addition, the interface between the assist device and the patient's blood requires careful consideration. Bernhard, LaFarge et al (32) favor artificial heart linings covered with either Dacron or velour; calcification may occur following prolonged pumping. Kolff favors the use of smooth polyurethane for the intima and has reported animals living for several months with an assist pump that utilized a polyurethane intima (28). Further work is necessary to determine the optimal intimal lining for the moving parts of the pump.

A variety of valves have been used in artificial hearts. Investigators currently favor glutaraldehyde-treated pig or cow heart valves. Hemolysis secondary to blood trauma has been diminished significantly in recent models of most pumps. The elimination of mechanical valves and appropriate design of pumping surfaces appears to have minimized blood trauma during cardiac assistance. Infection continues to be a limiting factor in implanted hearts with percutaneous drive lines. Velour coating of the drive lines is thought to reduce the risk of infection; however, this problem will not be eliminated until a totally implantable pumping device with an implantation engine and power source is developed.

The abdominal left ventricular assist device (ALVAD) was developed and extensively studied by Norman (33) and by Bernhard, LaFarge et al (32). Laboratory testing of a number of models of this type of assist device demonstrated that the pump may be left in the chest or abdomen for several months without untoward effects. The inlet tube, which is connected to the left ventricular apex, functioned well for over a year in the laboratory, and a modification of this tube has been used clinically with success in a number of patients (34). The ALVAD developed by Norman et al (35) consists of a pumping device implanted in the abdomen. Blood is removed from the apex of the left ventricle during systole and injected into the infrarenal portion of the abdominal aorta during diastole. The pump is of cylindrical design and operates axiosymmetrically. The pump ventricle is a polyurethane bladder that collapses when pneumatic pressure is applied to the space between the bladder and the stainless steel casing. Unidirectional flow is imparted by disc valves at the inflow and outflow orifices. In vitro testing of this pump shows that its output is a linear function of pumping rate, and the pump is effective while working synchronously, asynchronously, and even when the patient is in ventricular fibrillation (35). The pump is lined with "flocked" polyester fibrils, which rapidly develop a neointima consisting of a fibrin lattice. Positioning the pump in the abdomen results in two important advantages: pulmonary function is minimally impaired and removal may be accomplished through the abdomen without reentering the chest.

The ALVAD is currently entering a period of clinical trial in Texas. Initial experience in terminal patients confirmed the hemodynamic effectiveness of the ALVAD for supporting the circulation when the left ventricle failed. At the present time, the ALVAD appears to be the device that is most ready for clinical trial on a large scale. The ALVAD would be particularly useful in patients who are unable to be weaned from cardiopulmonary bypass despite maximal pharmacologic support and intraaortic balloon counterpulsation after corrective cardiac surgery. It is possible that the ALVAD will permit survival in patients whose recovery requires left ventricular support for a longer period of time than is currently available with present assist devices such as IABCP or left ventricular apex-to-aorta assistance. The ALVAD is capable of supporting the circulation of patients who do not respond to IABCP, and therefore expands the technology available to the surgeon attempting to treat end-stage myocardial failure after operation.

SUMMARY

We reviewed current concepts and techniques of cardiac assistance, including orthotopic and heterotopic cardiac allografting. Whereas intraaortic balloon counterpulsation has had the greatest clinical use, other techniques are now becoming available, including the implantable abdominal left ventricular assistance device, and much effort is being expended on the development of artificial hearts and their power sources.

Despite the present technical shortcomings, exciting advances have occurred in the clinical management of intractable acute or chronic ventricular failure and cardiogenic shock since the introduction of circulatory assistance in man. The continued development and implementation of these techniques will help reduce the still unacceptably high mortality rate in this group of patients.

Literature Cited

1. Stamler, J. 1974. In *The Myocardium: Failure and Infarction,* ed. E. Braunwald, Vol. 20, pp. 219–36. New York: H. P. Publishing. 409 pp.
2. Carrel, A., Guthrie, C. C. 1905. The transplantation of veins and organs. *Am. J. Med.* 10:1101–20
3. Lower, R., Shumway, N. 1960. Studies on orthotopic transplantation of the canine heart. *Surg. Forum* 11:18–20
4. Barnard, C. N. 1967. A human cardiac transplant: An interim report of a successful operation performed at Groote Schuur Hospital, Capetown. *S. Afr. Med. J.* 41:1271–74
5. Stinson, E. B. 1976. Appraising the status of heart transplantation. *Contemp. Surg.* 4:79–81
6. Griepp, R. B., Stinson, E. B., Bieber, C. P., Reitz, B. A., Copeland, J. G., Oyer, P. E., Shumway, N. E. 1976. Human heart transplantation: Current status. *Ann. Thorac. Surg.* 22:171–75
7. Caves, P. K., Stinson, E. B., Billingham, M. E., Shumway, N. E. 1974. Serial transvenous biopsy of the transplanted human heart. *Lancet* 1:821–26
8. Barnard, C. N., Losman, J. 1975. Left ventricular bypass. *S. Afr. Med. J.* 49:303–12
9. Gibbon, J. H. 1954. Application of a Mechanical Heart and Lung Apparatus to Cardiac Surgery. *Minn. Med.* 37:171–85
10. Hill, J. D., Gerbode, F. 1976. In *Gibbon's Surgery of the Chest,* ed. D. Sabiston, F. Spencer, Vol. 30, pp. 867–77. Philadelphia: Saunders. 1592 pp.
11. Bartlett, R., Jeffries, M., Haiduc, N., Gazzaniga, A. 1976. Extracorporeal membrane oxygenation cardiopulmonary support in infancy. *Am. Soc. Artif. Intern. Organs* 5:8 (Abstr.)
12. Salisbury, P. F., Bor, N., Lewin, R. J., Rieben, P. A. 1959. Effects of partial and total heart-lung bypass on the heart. *J. Appl. Physiol.* 14:458–69
13. Niell, W. A., Levine, H. J., Wagman, R. J., Gorlin, R. 1963. Left ventricular oxygen utilization in intact dogs: Effect of systemic hemodynamic factors. *Circ. Res.:* 163–72
14. Litwak, R. S., Koffsky, R. M., Lukban, S. B. 1974. Implanted heart assist device after intra-cardiac surgery. *N. Engl. J. Med.* 291:1341–48
15. Peters, J. L., McRea, J. C., Fukumasu, H., Kolff, W. J. 1976. Transapical left ventricular bypass (TALVB) without an auxiliary ventricle. *Trans. Am. Soc. Artif. Intern. Organs* 22:357–65
16. Kantrowitz, A., Kantrowitz, A. 1953. Experimental augmentation of coronary flow by retardation of the arterial pressure pulse. *Surgery* 34:678–87
17. Clauss, R. H., Birtwell, W. C., Albertol, G., Lunzer, S., Taylor, W. J., Fosberg, A., Harken, D. 1961. Assisted circulation I. The artificial counterpulsator. *J. Thorac. Cardiovasc. Surg.* 41:447–48
18. Moulopoulos, S. D., Topaz, S., Kolff, W. J. 1962. Diastolic pumping (with carbon dioxide) in the aorta—a mechanical assistance to the failing circulation. *Am. Heart J.* 63:669–75
19. Powell, W. J., Daggett, W. M., Magro, A. E., Bianco, J., Buckley, M., Sanders, C., Kantrowitz, A., Austen, W. G. 1970. Effects of intraaortic balloon counterpulsation on cardiac performance, oxygen consumption and coronary blood flow in dogs. *Circ. Res.* 26:753–64
20. Lamberti, J. J., Anagnostopoulos,

C. E., Al-Sadir, J., Das Gupta, D. S., Lin, C. Y., Replogle, R. L., Resenkov, L., Skinner, D. B. 1976. Mechanical circulatory assistance for the treatment of complications of coronary artery disease. *Surg. Clin. North Am.* 56:83–94

21. Lamberti, J. J., Cohn, L. H., Lesch, M., Collins, J. J. 1974. Intraaortic balloon counterpulsation for postoperative left ventricular power failure: Indications and long-term results. *Arch. Surg.* 109:766–71
22. Kaiser, G. C., Marco, J. P., Barner, H., Codd, J. E., Laks, H., William, V. 1976. Intra-aortic balloon assistance. *Ann. Thorac. Surg.* 21:487–91
23. Cohn, P. F., Gorlin, R., Herman, M. V., Sonnenblick, E., Horn, H., Cohn, L. Collins, J. J. 1975. Relation between contractile reserve and prognosis in patients with coronary artery disease and depressed ejection fraction. *Circulation* 51: 414–20
24. Dennis, C., Moreno, J. R., Hall, D. P., Grosz, C., Ross, S. M., Wesolowski, S., Senning, A. 1963. Studies on external counterpulsation as a potential measure for acute left heart failure. *Trans. Am. Soc. Artif. Intern. Organs* 9:186–91
25. Kantrowitz, A. 1976. Challenge to conventional treatment for myocardial failure— Mechanical assist. *Biomater. Med. Devices Artif. Organs 4:1–20*
26. Norman, J. C. 1975. The artificial heart: Perspectives, prospects, and problems of a high applied technology. *Bull. Tex. Heart. Inst.* 2:259–64
27. Akutzu, J. 1972. In *Cardiac Surgery,* ed. J. Norman, Vol. 33, pp. 671–77. New York: Appleton-Century-Crofts. 703 pp.
28. Kolff, W. J., Lawson, J. 1975. Status of the artificial heart and cardiac assist devices in the United States. *Trans. Am. Soc. Artif. Intern. Organs* 21:620–38
29. Schuder, J. C., Stephenson, H. E. Jr. 1963. Energy transport into the closed chest from a set of very large, mutually orthogonal coils. *Commun. Electr.* 64:527–35
30. Cooley, D. A., Liotta, D., Hallman, G. L., Bloodwell, R. D., Leachman, R. D., Milam, J. D. 1969. Orthotopic cardiac prosthesis for two staged cardiac replacement. *Am. J. Cardiol.* 24:723–30
31. Jarvik, R., Oster, H., Olsen, D., Volder, J., Stanley, R. 1974. Design of an elliptical ventricle and results of 26 implantations in calves. *Am. Soc. Artif. Intern. Organs* 3:29 (Abstr.)
32. Bernhard, W. F, Poirier, B. S., LaFarge, G., and Carr, J. 1975. A new method for temporary left-ventricular bypass. *J. Thorac. Cardiovasc. Surg.* 70:880–95
33. Norman, J. C. 1974. An abdominal left ventricular assist device (ALVAD)—perspectives and prospects. *Bull. Tex. Heart Inst.* 1:251–264
34. Cooley, D. A., Norman, J. C., Reul, G. C., Kidd, J., Nihill, M. R. 1976. Surgical treatment of left ventricular outflow tract obstruction with apico-aortic valved conduit. *Surgery* 80:674–80
35. Norman, J. C. 1976. An intracorporeal (abdominal) left ventricular assist device (ALVAD). XXX. Clinical readiness and initial trials in man. *Bull. Tex. Heart Inst.* 3:249–88

Ann. Rev. Med. 1978. 29:583–91
Copyright © 1978 by Annual Reviews Inc. All rights reserved

RABIES VACCINATION ◆7300

Stanley A. Plotkin, M.D. and Tadeusz Wiktor, D.V.M.
The Wistar Institute of Anatomy and Biology, Philadephia, Pennsylvania 19104

Vaccination against rabies carries an aura of mystery, even for physicians. In fact, some of this mystery is justified, for the mechanism by which a vaccine administered after introduction of the infecting agent can protect the host against disease has not been fully described. In addition, inoculation schedules with rabies vaccines have acquired the reputation of being painful and complicated, and vaccine failures ending in death have occurred with sufficient frequency to indicate the need for improved antigens. The objectives of this article are to review current vaccination policies and also to summarize studies concerning a new tissue culture vaccine, which may be licensed by the time this article is published.

PATHOPHYSIOLOGY OF VACCINE PREVENTION

It is probable that rabies virus remains latent, possibly in the muscle cells, for some time after an individual is exposed to the viral inoculum. As far as the action of vaccines is concerned, all that can be said with certainty is that antibody is preventive if it is present before exposure, and if vaccine is administered promptly after exposure, a state of resistance associated with antibody production usually prevents rabies from invading the central nervous system.

NERVE TISSUE VACCINE

Most of the rabies vaccine produced in the world today is of nerve tissue origin. Sheep brain vaccine and goat brain vaccine are used throughout large parts of Asia and Africa, but are associated with postvaccinal encephalitis (1). In South America and France, suckling mouse brain vaccines are widely used in the hope of depressing sensitization to nerve tissue, which causes encephalitis (2).

The efficacy of nerve tissue vaccine (NTV) in man has not been evaluated by controlled studies. The Pasteur Institute of Southern India analyzed the incidence of disease in vaccinated and unvaccinated subjects bitten by rabid animals. Untreated subjects developed rabies at the rate of 56%, compared with only 7% of vaccinated individuals (3).

0066-4219/78/0401-0583$01.00

On the other hand, NTV vaccination of persons bitten on the head or neck by rabid wolves in Iran was followed by a 40% overall mortality. That figure may be compared to a 47% rabies incidence (15 of 32) observed in one group of villagers who failed to seek treatment after being attacked by a rabid wolf (4).

Nerve tissue vaccines are essentially unavailable in the United States.

DUCK EMBRYO VACCINE

Duck embryo vaccine (DEV) was developed to overcome the problem of encephalitic reactions to NTV. In this respect it appears to have been successful; the incidence of neuroparalytic reactions was reduced from 1 in 1000 with NTV to 1 in 25,000 with DEV (5, 6). Unfortunately, a high price is paid in terms of immunogenicity. In postexposure treatment of Americans, Corey et al (7) found that 8% of those given courses of DEV failed to respond with the production of measurable antibody. Even more disturbing was the fact that 23% of recipients of DEV plus rabies antiserum did not develop antibody. The poor performance of DEV given prophylactically preexposure is described below.

In addition to the problem of antigenicity, DEV contains enough duck protein to produce a significant problem of allergic reaction. Rubin et al (5) performed a prospective study and found the toll of reactions reported in Table 1. Although not as deadly as neurologic complications, these reactions have limited the preexposure prophylactic use of DEV.

ANIMAL RABIES SERUM

Experimental evidence for the desirability of combining antiserum with vaccine was established in 1954 by Koprowski (8). The field test of combined vaccine-serum protection was made possible by a natural disaster, the attack of a rabid wolf on 29 Iranian villagers, 18 of whom were bitten on the head and neck (9). Rabies developed in three of five persons who were treated with NTV only, but in only 1 of 13 persons who were given antiserum with NTV. The single failure was in the subgroup of seven individuals who were given one dose of antiserum. Of the six patients who received two or more injections of serum with NTV, none succumbed.

Animal rabies serum (ARS) is produced in horses and dispensed in 10-ml vials containing 100 units/ml. The dose is 40 units/kg. Up to 500 units, or 50% of the

Table 1 Allergic reactions to duck embryo rabies vaccine[a]

Reactions	Preexposure vaccination (%)	Postexposure vaccination (%)
Local	67	100
Systemic	10	33
Anaphylaxis	0.5	0.9

[a] After Rubin et al (5).

calculated dose, should be infiltrated locally at the site of the wound, with the remainder given intramuscularly (i.m.).

HUMAN RABIES IMMUNE GLOBULIN

To eliminate the serum sickness reactions that follow ARS, volunteers were hyperimmunized with rabies vaccine and bled, and gamma globulin was prepared from their plasma. Human rabies immune globulin (HRIG) is now commercially available and, although expensive, will probably replace ARS for use in postexposure treatment. As a homologous antiserum, HRIG persists longer in the circulation of the host but may have an even greater dampening effect on active immunization. Thus Hattwick et al (10) found that 23 doses of DEV were needed to overcome the suppressive effect of 15–40 units of HRIG per kg.

The general recommendation for HRIG-DEV treatment is to give 20 IU of HRIG per kg immediately, with approximately half injected locally; then to administer 21 doses of DEV from days 0–14 (double doses for the first 7 days). Boosters are given on the 24th and 34th days after starting treatment.

HUMAN DIPLOID CELL VACCINE

The solution to the problems of antigenicity and safety of rabies vaccine evidently lay in the development of vaccines prepared from rabies virus grown in tissue culture free of neural tissue. Ideally the tissue culture would be of human cells to avoid foreign proteins. Several cell culture vaccines have been prepared from cultures of mammalian or avian cell origin. Hamster kidney rabies vaccine has been most thoroughly investigated starting with early work by Kissling (11). Immunogenic rabies vaccines have been prepared in hamster cell cultures, but problems of cell standardization and the possibility of sensitization have limited their general use.

Production of human diploid cell vaccine (HDCV) was made possible by the development of the WI-38 normal human fibroblast cell line by Hayflick & Moorhead (12) and the adaptation of the Pitman-Moore strain of rabies virus to growth in WI-38 by Wiktor et al (13). In practice, virus grown in WI-38 is concentrated by ultrafiltration to increase antigen content and then disaggregated and inactivated by tri-*n*-butyl phosphate (Wyeth) or inactivated by beta propiolactone (Merieux) (28).

EARLY TESTS FOR IMMUNOGENICITY

The earliest human tests were performed on the staff of the Wistar Institute and on veterinary students at the University of Pennsylvania (14, 15). Various schedules were tried, and it was found that three properly spaced i.m. doses invariably produced an immune response. Four doses administered within a 2-wk period seemed to give the fastest responses. As shown in Table 2, 1 ml injected i.m. at 0, 3, 7, and 14 days provided powerful immune responses to

Table 2 Antibody responses to first 4 lots of HDCV on schedule of 0, 3, 7, 14 days compared to DEV[a]

Lot No.	Level of Antibodies (I.U.)		
1	<0.1	3.2	2.5
2	<0.1	4.4	3.5
3	<0.1	4.3	7.4
4	<0.4	8.5	10.8
DEV	<0.1	0.14	0.13

[a] After Plotkin et al (15).

rabies antigen (15). A similar schedule of DEV resulted in titers 10- to 20-fold lower. Other workers also showed excellent immunogenicity, both by the i.m. and intradermal routes (16–20).

LARGE-SCALE TESTS FOR IMMUNOGENICITY IN PREEXPOSURE USE

It was natural to turn to veterinary students for larger tests of HDCV, since many of them were receiving prophylactic DEV routinely and were anxious for better protection. The schools where vaccination was performed and the approximate numbers vaccinated are listed in Table 3. A schedule of 1 ml i.m. at 0, 7, and 21 days was used.

Table 3 Number of veterinary students and other volunteers receiving primary pre-exposure vaccinations with HDCV

Institution	Students (No.)
Auburn	76
California	81
CDC	80
Colorado	96
Cornell	72
Iowa	204
Louisiana	69
Michigan	86
Minnesota	96
Missouri	103
Ohio	130
Oklahoma	80
Oregon	6
Pennsylvania	258
Texas	246
Wistar	89
Total	1772

Physicians observed vaccinees at each school and recorded vaccination data. However, since there were different observers and different methods of reporting reactions and no placebo group, it is difficult to state the precise reaction rates. In students at the University of Pennsylvania reactions to HDCV were generally less significant, and certainly no more severe, than those of Auburn students inoculated with DEV (21). The actual data for the University of California are given in Table 4. The variety of nonspecific complaints is apparent. An overall estimate of reaction rates is given in Table 5.

Rabies antibody titers were determined on the serum of each veterinary student. Virtually 100% of volunteers developed antibodies, with geometric mean titers at each school of 10 IU or greater, which in our laboratory is equivalent to titers of at least 1/400. A direct comparison was made with titers of DEV administered at Auburn by various schedules (21). The differences in results after schedules of 0, 7, and 21 days or after 0, 30, and 180 days (the presently suggested schedule for DEV) are stark and need no emphasis (Table 6).

Table 4 Reactions following vaccination with HDCV (University of California—81 vaccinees)

Reaction	1st Dose	2nd Dose	3rd Dose
Sore arm	41	38	27
Headache	3	3	2
Nausea	3	2	1
Fatigue	6	1	3
Dizziness	2	1	0
Deafness (transitory)	1	0	0
Thirst	1	1	0
Malaise	2	0	0
Drowsy	1	0	0
Insomnia	1	1	1
Jittery	0	0	1
Diarrhea	1	0	0
Rash	1	0	0
Fever	1	0	0
No reaction	38	43	48
Local reaction	41	38	27
Systemic reaction	6	3	3

Table 5 Reaction rates observed in over 1770 human volunteers receiving HDCV

Reaction	Rate (%)
Sore arm	15–25
Headache	5–8
Malaise and/or nausea	2–5
Allergic edema	0.1

Table 6 Antibody responses to HDCV or DEV (administered at Auburn University)

	Schedule			
	0, 7, 21 days		0, 30, 180 days	
Titers (I.U.)	HDCV	DEV	HDCV(2)	DEV
<0.5	—	19	—	32
0.5–2.5	1	13	11	17
2.6–10.0	2	3	16	1
≥10.0	36	1	9	—

PERSISTENCE AND BOOSTERS

Even with HDCV, antibodies fall off rapidly after initial immunization. However, profound differences exist between the situation after HDCV and that after DEV.

First, most HDCV vaccinees have some detectable antibody at two years after initial vaccination, in contrast to DEV vaccinees, most of whom are negative at one year. Second, one booster of HDCV given to previously HDCV-vaccinated individuals results in a dramatic anamnestic response, with titers rising from 2.8 units to 94 units at 14 days and more than 100 units in 35 days. The same booster given to DEV vaccinees results in a more modest booster effect. It should be noted that only those who responded to DEV at some time in the past will respond to a booster; individuals who never responded to DEV with any antibody are not sensitized and need a complete schedule of immunization with HDCV.

POSTEXPOSURE VACCINATION

The crucial test of a rabies vaccine is protection of those actually exposed to the virus. Throughout the world most vaccinations are performed for possible exposure, and not for exposure confirmed by the demonstration of virus in the brain of the biting animal. Collection of cases of proven rabies exposure is therefore not easy, and since a controlled trial of vaccine is not ethically permissible, we must depend on observation of vaccinees after proven exposure and the comparison of these results to previous experience with unvaccinated individuals.

HDCV has been used postexposure primarily in West Germany, Iran, France, and the USA. In the latter two countries a total of at least 500 individuals have been vaccinated "postexposure"; however, in relatively few cases has there been proof of rabies in the biting animal. In Germany and Iran, however, the use of HDCV has been backed up by unequivocal laboratory data that so far has shown HDCV to have outstanding efficiency.

Rabies in foxes has been steadily moving westward in Europe, and Germany is heavily involved in this epizootic. Kuwert (22), in Essen, vaccinated 44 individuals after proven exposure to rabid dogs, cats, cows, and wild animals, or to rabies virus in the laboratory. The schedule used was 1 ml i.m. on days, 0, 3, 7, 14, 30, and 90. There were no failures of protection, no significant reactions, and excellent neutralizing and complement-fixing antibody responses.

An even more dramatic test of HDCV was conducted by Bahmanyar et al (23) in Iran. Forty-five individuals who were bitten by rabid wolves or dogs were given rabies antiserum followed by the same schedule of vaccine as that used in West Germany. Once again, no rabies was seen in vaccinees, despite a 40% risk (estimated from previous experience) if they had remained unvaccinated. Antibody measurements showed mean titers as follows: 7 days, 1.1 IU; 14 days, 10.7 IU; 30 days, 49 IU; and 100 days, 321 IU.

The Center for Disease Control (CDC) has been distributing HDCV of American manufacture postexposure to patients allergic to DEV or failing to respond to DEV. Approximately 80 such people have been immunized, but it must be noted that risk of rabies and the extent of previous DEV administered were quite variable. Nevertheless, no vaccine failures have occurred to date, and all who received the full schedule of four 1-ml doses i.m. on days, 0, 7, 14, and 21 have responded with antibodies.

LOCAL TREATMENT

Removal of rabies virus–containing saliva is a crucial part of treatment and should be undertaken urgently. The wound should be flushed copiously with soap and water, and ethyl alcohol, in whatever form is immediately available, should then be applied locally. The concentration of ethanol is important; 43% (86 proof) or higher gives the best results (24).

Later, the wound should be flushed again with 20% soap solution or 1–2% benzalkonium chloride (25). Note that the latter is used in a much higher than usual concentration. Substitutes for benzalkonium (Zephiran ®) include cetrimonium bromide (Cetavlon ®) or Hyamine ®. The application of fuming nitric acid is no longer considered necessary. Experimental data suggest that regardless of the solution used, adequate flushing is important, particularly in puncture wounds. Catheters should be inserted into puncture wounds and fluid instilled by means of an attached syringe (26, 27). If this proves too painful, the area can be anesthetized safely with local procaine-type anesthetics (24, 25).

GENERAL RECOMMENDATIONS

Table 7 summarizes the decisions that must be made concerning whether or not to give rabies vaccine. As in all other areas of medicine, each case must be considered individually, and the help of rabies experts should be freely solicited.

Schedules for postexposure use are summarized in Table 8. The schedule recommended by the author is based on the slightly better antibody response on day 7 if a dose is given on day 3, and the need for a booster dose (day 28) to overcome the effect of HRIG. However, the CDC schedule seems to have been quite effective in practice with regard to antibody response. On the other hand, the World Health Organization (WHO) schedule, which the authors had a hand in formulating, has survived the test of severe rabies challenge in Germany and Iran.

Table 7 Postexposure antirabies treatment guide[a]

Animal	Evaluation of animal at time of exposure[b]	Treatment of exposed human
Wild		
Skunk Fox Raccoon Coyote Bat	Regard as rabid	HRIG + rabies vaccine[c]
Domestic		
Dogs Cats	Healthy	None[d]
	Escaped (unknown)	HRIG + rabies vaccine
	Rabid or suspect rabid	HRIG + rabies vaccine[c]

[a] Modified from Public Health Service Advisory Committee Recommendations. 1977. *Ann. Intern. Med.* 86:452–55. These recommendations are only a guide. They should be used in conjunction with knowledge of the animal species involved, circumstances of the bite or other exposure, vaccination status of the animal, and presence of rabies in the region.

[b] An exposure is considered to be by bite, by scratch with claws, or by contamination with saliva of mucosal surfaces or skin that has been cut or abraded.

[c] Discontinue vaccine if fluorescent antibody tests of animal are negative.

[d] Begin HRIG + rabies vaccine at first sign of rabies in biting dog or cat during holding period (10 days).

Table 8 Possible schedules for postexposure use of HDCV

Schedule of doses (days)	Route	Recommended by
0, 3, 7, 14, 30, 90	i.m.	WHO Expert Committee
0, 7, 14, 21	i.m.	CDC
0, 3, 7, 14, 28	i.m.	Author, for rapid antibody response and high titer
0.1 ml in 4 sites on day 0	i.d.[a]	Turner et al (19) for rapid antibody response

[a] Intradermal.

Literature Cited

1. Applebaum, E., Greenberg, M., Nelson, J. 1953. Neurological complications following antirabies vaccination. *J. Am. Med. Assoc.* 151: 188–91
2. Fuenzalida, E., Palacios, R., Borgono, J. M. 1964. Antirabies antibody response in man to vaccine made from infected suckling-mouse brains. *Bull. WHO* 30:431–36
3. Veerarghavan, N. 1969. *Sci. Rep. 1968 Pasteur Inst. South.* India. pp. 35–42
4. Hildreth, E. A. 1963. Prevention of

rabies, or the decline of Sirius. *Intern. Med.* 58:883–96

5. Rubin, R. H., Hattwick, M. A. W., Jones, S., Gregg, M. B., Schwartz, V. D. 1973 Adverse reactions to duck embryo rabies vaccine. *Ann. Intern. Med.* 78:643–49
6. Plotkin, S. A., Clark H. F. 1971. Committee on Immunization: Prevention of rabies in man. *J. Infect. Dis.* 123:227–40
7. Corey, L., Hattwick, M. A. W., Baer, G. M., Smith, J. S. 1976. Serum neutralizing antibody after rabies postexposure prophylaxis. *Ann. Intern. Med.* 85:170–76
8. Koprowski, H., Black, J. 1954. Studies on chick-embryo-adapted rabies virus. V. Protection of animals with antiserum and living attenuated virus after exposure to street strain of rabies virus. *J. Immunol.* 72:85–93
9. Baltazard, M., Bahmanyar, M., Ghodssi, M., Sabeti, A., Gajdusek, C., Rouzbehi, E. 1955. Essai pratique du serum antirabique les mordus par loups enrages. *Bull. WHO* 13:747–72
10. Hattwick, M. A. W., Rubin, R. H., Music, S., Sikes, R. K., Smith, J. S., Gregg, M. B. 1974. Postexposure rabies prophylaxis with human rabies immune globulin. *J. Am. Med. Assoc.* 227:407–10
11. Kissling, R. E., Reese, D. R. 1963. Anti-rabies vaccine of tissue culture origin. *J. Immunol.* 91:362–68
12. Hayflick, L., Moorhead, P. S. 1961. The serial cultivation of human diploid cell strains. *Exp. Cell. Res.* 25:585–621
13. Wiktor, T. J., Fernandes, M. V., Koprowski, H. 1964. Cultivation of rabies virus in human diploid cell strain, WI-38. *J. Immunol.* 93:353–66
14. Wiktor, T. J., Plotkin, S. A., Grella, D. W. 1973. Human cell culture rabies vaccine. *J. Am. Med. Assoc.* 224:1170–71
15. Plotkin, S. A., Wiktor, T. J., Koprowski, H., Rosanoff, E. I., Tint, H. 1976 Immunization schedules for the new human diploid cell vaccine against rabies. *Am. J. Epidemiol.* 103:75–80
16. Cabasso, V. J., Dobkin, M. B., Roby, R. E., Hammar, A. H. 1974. Antibody response to a human diploid cell rabies vaccine. *Appl. Microbiol.* 27:553–61
17. Aoki, F. Y., Tyrell, D. A. J., Hill, L. E., Turner, G. S. 1975. Immunogenicity and acceptability of a human diploid-cell culture rabies vaccine in volunteers. *Lancet* 1:660–62
18. Cox, J. H., Schneider, L. G. 1976. Prophylactic immunization of humans against rabies by intradermal inoculation of human diploid cell culture vaccine. *J. Clin. Microbiol.* 3:96–101
19. Turner, G. S., Aoki, F. Y., Nicholson, K. G., Tyrell, D. A. J., Hill, L. E. 1976. Human diploid cell strain rabies vaccine: Rapid prophylactic immunisation of volunteers with small doses. *Lancet* 1:1379–81
20. Garner, W. R., Jones, D. O., Pratt, E. 1976. Problems associated with rabies pre-exposure prophylaxis. *J. Am. Med. Assoc.* 235:1131–32
21. Kramer, T. T. Personal communication
22. Kuwert, E. K., Marcus, I., Hoher, P. G. 1976. Neutralizing and complement-fixing antibody responses in pre- and postexposure vaccinees to a rabies vaccine produced in human diploid cells. *J. Biol. Stand.* 4:249–62
23. Bahmanyar, M., Fayaz, A., Nour-Salehi, S., Mohammadi, M., Koprowski, H. 1976. Successful protection of humans exposed to rabies infection: postexposure treatment with the new human diploid cell rabies vaccine and antirabies serum. *J. Am. Med. Assoc.* 236:2751–54
24. Wiktor, T. J., Koprowski, H. 1963. Action locale de certains medicaments sur l'infection rabique de la souris. *Bull WHO* 28:487–94
25. Kaplan, M. M., Cohen, D., Koprowski, H., Dean, D., Ferrigan, L. 1962. Studies on the local treatment of wounds for the prevention of rabies. *Bull WHO* 26:765–75
26. Dean, D. J., Baer, G. M., Thompson, W. R. 1963. Studies on the local treatment of rabies-infected wounds. *Bull WHO* 28:477–86
27. Shaughnessy, H. J., Zichis, J. 1954. Treatment of wounds inflicted by rabid animals. *Bull WHO* 10:805–13
28. Tint, H., Dobkin, M. B., Rubin, B. A. 1974. A new tissue culture (WI-38) rabies vaccine, inactivated and disaggregated with tri-(n)-butyl phosphate. Int. Symp. Rabies, Lyon 1972, *Symp. Ser. Immunobiol. Standard* 21:132–44

Ann. Rev. Med. 1978. 29:593–603
Copyright © 1978 by Annual Reviews Inc. All rights reserved

RECENT MEMBRANE RESEARCH AND ITS IMPLICATIONS FOR CLINICAL MEDICINE

◆7301

Vincent T. Marchesi, M.D., Ph.D.
Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06510

LIPIDS ARE ARRANGED AS A DYNAMIC ASYMMETRIC BILAYER

One of the most remarkable structural features of biological membranes was first clearly described in the mid 1920s when Gorter & Grendel (38) suggested that the lipids of the red cell membrane might be arranged in the form of a double layer of phospholipid molecules. This remarkably simple idea was based on the fact that the surface area of red cell ghosts contained approximately twice as many lipid molecules as were needed to make a monomolecular film surrounding the cell. This lipid-bilayer hypothesis received a certain amount of support from early X-ray diffraction studies carried out on model membrane systems and on myelin membranes, and lately it has been supported by studies using more modern physical techniques. The development of freeze-etching techniques was a significant step forward in the analysis of membranes by electron microscopy. With this approach it was discovered that cell membranes cleaved roughly in half along planes that corresponded to the middle of the hydrocarbon regions of the lipid bilayers (1).

Studies with some of the newer physical techniques also revealed that lipid molecules in bilayers appeared to be in very active motion. By placing probes on different segments of phospholipid molecules and sampling different layers of the hydrocarbon matrix, it has been shown that the hydrocarbon chains of fatty acids of the phospholipids are rapidly flexing. Individual phospholipid molecules also move at significant rates within the lateral plane of the lipid bilayer. Hence, it is now recognized that the individual phospholipid molecules are in a dynamic state within the intact membrane (2).

These new discoveries resulted in the concept of membrane fluidity. In general terms, fluidity refers to the degree of viscosity objects would encounter if they were able to float freely within the interior of the bilayer. The fluidity of a

0066-4219/78/0401-0593$01.00

membrane is determined by the types of lipids and fatty acids that make up the bilayer and by the temperature of the system. As rough generalizations, it may be stated that fatty acids with greater degrees of unsaturation are more fluid at lower temperatures than fatty acids that are more fully saturated. Also, shorter-chain fatty acids are more fluid than longer-chain fatty acids. Recent studies on the role of sterols and their effects on membrane fluidity are interesting in that sterols have a somewhat paradoxical effect: They can either increase or decrease the relative fluidity of membranes, depending upon the types of phospholipids and the concentrations of sterols in the preparations.

One additional recent modification of the bilayer hypothesis is the fact that the individual classes of lipid molecules appear to be arranged asymmetrically within the bilayers of intact cells. In the case of red blood cells, phosphatidylcholine and sphingomyelin, which are choline-containing phospholipids, seem to be concentrated in the external leaflet of the bilayer, while phosphatidylserine and phosphatidylethanolamine are concentrated in the cytoplasmic half (3). Now that the concepts of membrane fluidity and lipid asymmetry are more or less established, investigators are actively pursuing the biological meaning of these remarkable properties.

INTEGRAL PROTEINS INTERACT WITH LIPIDS VIA HYDROPHOBIC ASSOCIATIONS; EXTRINSIC PROTEINS (MAY) INTERACT ELECTROSTATICALLY

Approximately one half of the mass of most mammalian cell membrane fractions is made up of proteins, and the number of types of different polypeptide chains seems to vary with the cell type. Human red blood cells have at least seven major polypeptide chains, as determined by sodium dodecyl sulfate (SDS) gel electrophoresis (Figure 1), and in addition there are an indeterminate number of minor protein species.

As a result of the initial attempts to extract proteins from cell membrane fractions, it became evident that some of the polypeptide chains were easily detached from membrane elements, while others seemed resistant to all but the most disruptive solubilizing agents. On the basis of such experiences from many different laboratories, Singer (5) suggested that polypeptide chains of red cell membranes might be operationally classified into two types, based on the apparent mode by which they associate with the membrane lipids. He suggested that the polypeptide chains that are relatively loosely bound to the membrane and can be removed by manipulating the pH or ionic strength of solubilizing media be called extrinsic membrane proteins. One of the most characteristic of the extrinsic membrane proteins of the red cell membrane is the set of polypeptide chains called spectrin, which appears as bands one and two on SDS gels.

Many investigators discovered that the spectrin polypeptides could be rapidly and quantitatively eluted from red cell ghosts by simply immersing the membranes in low-ionic-strength media that contained a dilute chelating agent (6). Although red cell membranes treated this way fragment into small vesicles as a

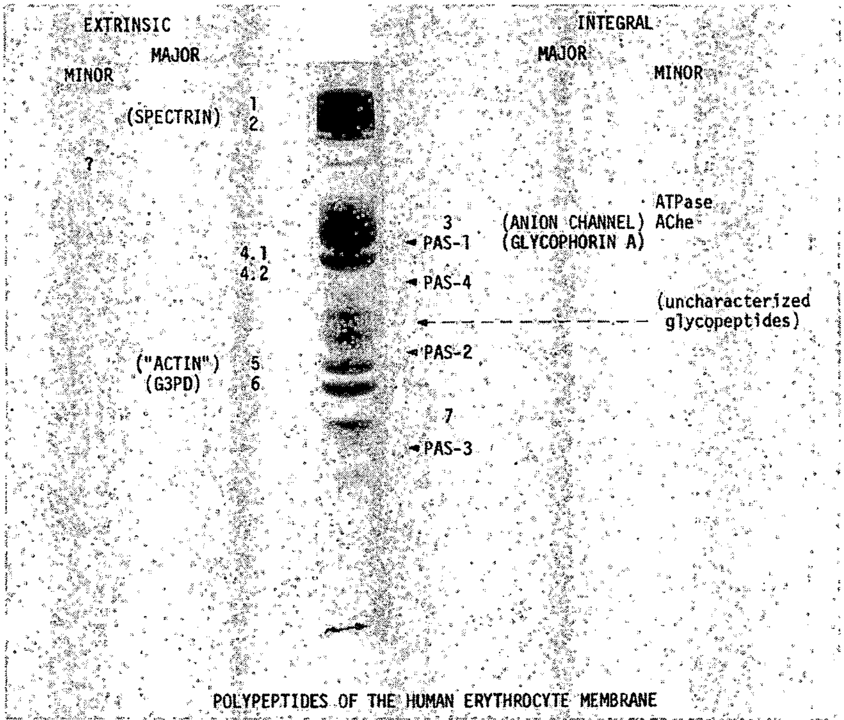


Figure 1 Electropherogram of the proteins of human red cell membranes in SDS polyacrylamide gels detected by staining with Coomassie brilliant blue. The numerical designation for the major polypeptides is adopted from Fairbanks et al (4). The bands indicated as 4.1 and 4.2 are not separated under the conditions used (1% SDS). For the sake of simplicity, designations for minor components are not included. The positions where the four sialoglycopeptides migrate on appropriately stained gels (PAS) are indicated by arrows.

result of the spectrin released from the membrane, it is generally thought that the spectrin polypeptides probably do not intercalate within the lipid bilayer, but instead associate electrostatically either with other proteins of the membrane or, possibly, with the polar groups of the phospholipids.

In contrast to extrinsic membrane proteins, the polypeptide chains that resist solubilization from membranes by simple buffer washings represent an entirely different class of macromolecules. These proteins are called integral proteins, largely because they cannot be dissociated from the lipid elements of the membrane without the use of detergents or harsh denaturing agents. Hence, it is considered likely that these proteins form an integral part of the membrane structure. One of the most characteristic of the integral membrane proteins is the major sialoglycoprotein of the human red cell membrane called glycophorin A.

Glycophorin A has been isolated by a number of different laboratories using a wide variety of protein solvents (see 7–9), and as the result of an intensive amount of study, it is now possible to construct a provisional model of its primary structure and offer some ideas of how and where it is oriented in the intact red cell membrane. The complete amino acid sequence and the sites of oligosaccharide attachment were recently determined (10), and are shown in Figure 2.

The polypeptide portion of glycophorin A, which comprises approximately 40% of the total dry weight, is made up of 131 amino acids. The distribution of different amino acids is rather interesting in that there is a very high concentration of threonine and serine residues located near the N-terminal end of the polypeptide chain. There is also a striking concentration of nonpolar amino acids located approximately midway between the N-terminal third of the polypeptide and the C-terminal third. In the diagram shown in Figure 2, the segment starting from amino acid 71 and extending through amino acid 90 is depicted between two vertical lines; this is thought to be the segment of polypeptide that is buried within the lipid bilayer.

It is also evident from this diagram that there are 16 oligosaccharide chains attached to the polypeptide chain, and their positions are assigned to the threonine and serine residues shown. It is significant that the glycosylated residues start from the N-terminal end of the molecule and extend to residue 50, but no additional sugar residues have been detected in amino acids beyond residue 50. One of the oligosaccharide chains linked to asparagine is of the complex variety (11), while the remaining 15 seem to be of the tetrasaccharide type originally described by Winzler & Thomas (12).

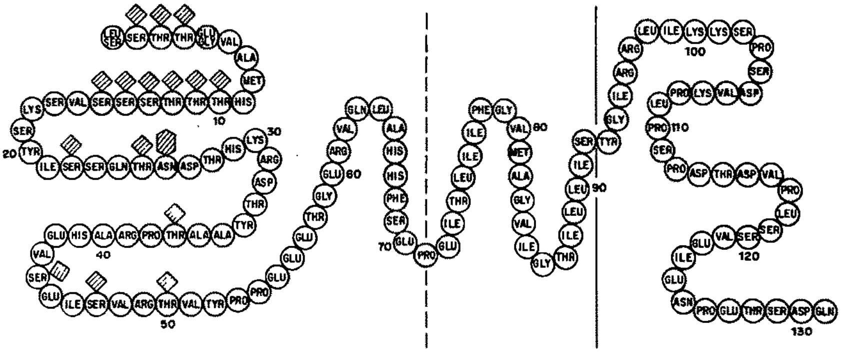


Figure 2 The amino acids of glycophorin A are arranged in this diagram to simulate, in a very general way, the positions they might have if the glycophorin molecule were arranged perpendicular to the lipid bilayer of the membrane. The limits of the bilayer are defined by the two vertical lines. The solid vertical line, which passes between residues 92 and 93, should be the approximate location of the polar groups of the inner half of the phospholipid bilayer. This assignment is based on the results of enzymatic iodination of tyrosine 93 and the distribution of ferritin-antibody conjugates directed against antigenic determinants defined by residues 102 to 118. Since we do not have comparable data with regard to the amount of the N-terminal end of glycophorin A that is buried with the lipid bilayer, we can only guess at the location of this outer lamella of the bilayer relative to the glycophorin molecule. Hence, the outer edge of the bilayer is defined by the dashed vertical line.

Since all of the glycosylated residues are concentrated at one end of the molecule, and some of these sugars probably represent receptor sites for lectins and perhaps for some blood groups antigens, this end of the molecule has been designated as the receptor domain.

The amino acid sequence of the 40 C-terminal amino acids is unusual in that this portion of the polypeptide chain contains a large number of charged amino acids with a peculiar clustering of acidic amino acids at the C-terminal segment of the chain. This portion of the polypeptide chain also contains a substantial number of prolines, which may play a major role in determining the conformation of this part of the molecule.

The peptide stretch of 22 amino acids, which connects the glycosylated receptor region with the C-terminal end (Figure 2), is composed solely of nonpolar residues, and it is logical to suggest that this segment is the part of glycophorin that interacts with the lipids of the membrane.

Some Integral Membrane Proteins Span the Lipid Bilayer

There is a considerable amount of experimental data in support of the idea that glycophorin A, and other integral membrane proteins of the red cell, are in a transmembrane configuration. This idea was originally based on the attempts to differentially label portions of the polypeptide chain in intact cell membranes and from the membrane preparations that resulted from osmotic lysis. Bretscher (13) originally showed that a segment of the polypeptide chain of glycophorin A was not accessible to labeling by a radioactive isotope introduced outside intact cells; however, this inaccessible segment could be labeled when the cell membranes were damaged or rendered permeable to the reagent. One interpretation of this experiment was that part of the polypeptide chain of the molecule extends outside the intact red cell, and another part is either buried within the membrane or situated in the cytoplasmic compartment of the cell. This experiment was repeated in several laboratories using a gentler labeling procedure, lactoperoxidase iodination.

In theory, lactoperoxidase iodination should be an effective probe for analyzing the orientation of glycophorin A, since this molecule has four tyrosines, three of which are located in the glycosylated portion of the polypeptide chain, while the fourth is close to the C-terminal end. Thus, three tyrosines should be labeled when intact cells are incubated with lactoperoxidase, while the fourth, located at position 93, should not be labeled unless the permeability of the membrane is broken. Several investigators have reported the expected results (14, 15), although other investigators were not able to reproduce these findings (16, 17).

In the light of the controversy surrounding the results of radiolabeling, we recently reinvestigated this problem and attempted to study the orientation of glycophorin A in intact red cells using ferritin-conjugated antibodies directed against specific polypeptide segments of glycophorin A.

Antibodies prepared against a 17-amino-acid peptide derived from the C-terminal end of this molecule localized uniformly along the inner surfaces of intact red blood cells (18). This was achieved by incubating frozen sections of glutaraldehyde-fixed red cells with ferritin-conjugated antibodies.

Comparable radiolabeling studies have shown that the band 3 protein also assumes a transmembrane orientation while in the red cell membrane. The fact that there are two major integral proteins of the red cell membrane and both have a transmembrane orientation suggests that perhaps all integral membrane proteins are transmembrane. This orientation might be the result of some special biosynthetic mechanisms responsible for membrane protein synthesis or it might be related to the functions of integral proteins. However, we cannot draw any conclusions until we find out whether integral membrane proteins of other cells are also transmembrane. It will also be important for us to have a better idea of how a transmembrane orientation contributes to the functional properties of these molecules.

Membrane Glycoproteins May Exist as Multimeric Complexes

Recent studies suggest that some membrane glycoproteins have a remarkable capacity to form multimeric complexes in the presence of sodium dodecyl sulfate (SDS) (19, 20). Almost all glycophorin preparations analyzed by SDS acrylamide gels contain multiple PAS-staining bands, usually designated PAS-1, -2, and -3. These bands correspond to molecular weights of 83,000, 45,000, and 25,000 when they are compared with the mobilities of standard proteins. However, we are not sure that these values represent the true molecular weights of the glycoproteins, since they are known to migrate anomalously on SDS gels (21). PAS-1 is the predominant form and represents approximately 75% of the total.

When this pattern was first obtained, it was not clear whether the multiple bands represented a series of chemically distinct polypeptide chains, or whether the smaller PAS-staining bands represented degradation products of the larger component. In a series of recent studies we showed that two of the PAS-staining bands (PAS-1 and PAS-2) were interconvertible, and their relative amounts depended on the conditions used to prepare the glycoproteins for SDS electrophoresis. Glycophorin A appears to exist as a dimeric form, which corresponds to PAS-1, and this dimeric form seems to be stabilized by noncovalent associations between hydrophobic segments of its polypeptide chains (20). This association can be disrupted by heating the glycoproteins in SDS or by modifying the methionine located in position 81 by specific alkylation in the presence of denaturing agents (22).

The conditions for the selective alkylation are extremely stringent: High concentrations of alkylating agents must be used in the presence of either urea or guanidine, and the glycoprotein must have been previously delipidated with organic solvents. Since the presence of either lipid or SDS interferes with the alkylation of methionine 81, and in so doing renders the molecular aggregate less susceptible to dissociation, it is conceivable that when the molecule is oriented in the intact cell membrane, the surrounding membrane lipids serve a similar function. On the basis of this, we might speculate that glycophorin A actually exists as dimeric, or possibly, even multimeric forms in situ. If this proves to be true, then it is possible for multimer-monomer association and dissociation reactions to take place in membranes, which could play a role in regulating the functions of these macromolecules. This hypothetical idea is consistent with recent studies on the apparent mobility of membrane components and the important role the lipid matrix may play in regulating this process.

Band 3 also seems to exist in the form of a dimer held together by noncovalent associations (29). Band 3 has a particularly reactive sulfhydryl group on its cytoplasmic segment, which can be oxidized to form band 3 dimers held together by a disulfide linkage (29). It is conceivable that the oligomeric state of band 3 can be influenced by the oxidation-reduction conditions inside the cytoplasm of the intact red cell.

The ability of membrane glycoproteins to aggregate reversibly could have great physiologic implications, if the molecules function as receptors for hormones and other ligands. The formation (or dissociation) of oligomers could serve as an on-off switching mechanism, if, for example, hormone activation required some specific multimeric form. Similarly, changes in the aggregation state of membrane glycoproteins must also have marked effects on the antigenic properties of the external position of such molecules. Neoantigens could be created by the close proximity of the receptor domains of glycoproteins brought together in the process of creating multimeric forms. The potential for creating a large number of antigenic forms exists, if different glycoproteins are able to form heteroaggregates. Clearly the factors that regulate the formation of glycoprotein complexes must be understood before we can begin to approach these questions.

MEMBRANE PROTEINS ARE DISTRIBUTED IN A MOSAIC: SOME MAY BE MORE MOBILE THAN OTHERS

One of the most intriguing and still largely unanswered questions in membrane biology concerns the three-dimensional arrangement of proteins at the cell surface. Do membrane proteins occupy fixed positions relative to some internal cytoplasmic structure, or are they in continuous motion, constantly changing their positions within the bilayer? One of the earliest suggestions concerning the arrangement of membrane proteins was that monomolecular sheets of polypeptide coated the phospholipid polar groups of the lipid bilayer. For many reasons, this idea now seems unlikely. Most of the evidence suggests that the bulk of the red cell membrane proteins exist as discrete macromolecular complexes, which are attached to the lipid bilayer by insertion of hydrophobic segments into the hydrocarbon interior (like glycophorin A) or are bound to polar groups of lipids electrostatically.

Our conceptions about the arrangement of proteins in membranes have also been heavily influenced by the spectacular experiments demonstrating membrane fluidity and phospholipid mobility described above. Since the lipid matrix in which the protein molecules must be inserted is in a constant state of flux, many have reasoned that protein molecules are probably also in a mobile state, although it is not clear whether they are simply floating passively in the lipid sea or are directed by some other agents independent of the surrounding lipids.

A vivid demonstration of the fact that some membrane proteins can be mobilized by external forces has been provided by the results of applying multivalent ligands to lymphocytes and other cells in suspension. Appropriately labeled ligands form patches and caps on such cells within minutes after their application. Other evidence for the mobility of surface components comes from cell fusion studies that show that surface antigens can freely intermix over the combined surfaces of hybrid cells at a remarkably rapid rate (23). Both sets of observations demonstrate rather unequivocally that certain types of surface macromolecules are capable of moving within the plane of the membrane. Unfortunately, the dramatic nature of these experiments may be overemphasizing the potential importance of the mobility of membrane proteins. It is likely that only certain types of membrane proteins are completely mobile, while others may be more or less in a fixed position at the cell surface relative to some cytoplasmic structures. We still do not know whether any proteins in the human red cell membrane are potentially mobile, or whether all of the major integral membrane proteins are firmly fixed at their respective sites.

Two of the most prominent integral proteins of the red cell membrane, band 3 and glycophorin A, probably are associated with the intramembranous particles seen by freeze-etching (24, 25). Since these particles do not appear to move within the plane of the intact membrane (based on their inability to patch or cap by the usual techniques), it is reasonable to conclude that these proteins are not extremely mobile. The intramembranous particles are movable, however, when osmotically lysed membranes are incubated with protease (24) or pH-5.5 buffers (26).

Integral Membrane Protein Interaction with Other Membrane Components

There have been several recent provocative studies purporting to show that glycophorin A and band 3 may have specific interactions with other membrane proteins. In particular, it has been suggested that the buried or cytoplasmic segment of glycophorin A may have the capacity to interact with the spectrin polymers, and the resulting complex may have some important function in regulating the stability of the membrane or in determining the relative positions of different membrane proteins. This exciting idea is based on the capacity of antispectrin antibodies to modify the topographical distribution of glycophorin A molecules (27), and on recent attempts to modify the distribution of intramembranous complexes by manipulating the polymeric state of spectrin.

Branton and co-workers (26, 28, 30) confirmed that incubation of red blood cell ghost membranes at pH 5.5 caused the reversible aggregation of the intramembranous particles, but they also showed, that this effect only occurred if the ghost membranes were partially depleted of their spectrin. They reason that spectrin normally restrains particle mobility, and shifts occur in spectrin-depleted membranes at the isoelectric point of spectrin because the residual aggregated spectrin can rearrange the intramembranous particles without opposition.

Yu & Branton (31) also showed that artificial reconstituted liposomes containing band 3, spectrin, and band 5 (actin) could be formed, and intramembranous particles would aggregate when the liposomes were incubated at pH 5.5. This remarkable result is the first demonstration that the components involved in the formation and regulation of the intramembranous particles can be dissociated and reconstituted into chemically defined liposomes. The authors note, however, that the components used to prepare the reconstituted vesicles were not completely characterized, and thus they do not rule out the participation of other proteins in the binding of spectrin to the membranes. Bennett & Branton (32) recently described conditions for the binding of purified spectrin back onto the inner surfaces of red blood cell ghost membranes through use of inside-out vesicles, and they found that neither glycophorin A nor band 3 were essential for this process.

Many investigators have found that the shape and deformability of the red cell membrane is related in some as yet undefined way to the ATP and calcium content of the cell (33–35). Sheetz & Singer (36) found that red blood cell ghosts underwent a shape change in the presence of salt solutions. This change required ATP and was enhanced by antispectrin antibodies. They attribute this effect to the crosslinking or polymerization of spectrin that is coupled to the hydrolysis of ATP. The process is enhanced by antispectrin antibodies as the latter further promote spectrin-spectrin associations.

Birchmeier & Singer (37) also found that the ATP utilized during the shape change of the red blood cell membranes was involved in the phosphorylation of band 2 of spectrin rather than serving as a substrate for a myosin-like ATPase. These investigators showed that incorporation of phosphatases inside red blood cell ghost membranes during osmotic lysis caused a reversal of the shape change effected by incubating the membranes with ATP.

These recent studies provide convincing evidence that spectrin plays a major role in regulating the shape and macromolecular topography of the red cell membrane. It is also clear that further insight into the mechanisms behind these processes, and their physiological and pathological implications, will only be achieved when we have a much clearer idea of the molecular properties of each of the interacting units.

SUMMARY

Our understanding of the molecular organization of even the simplest mammalian cell membrane—that of the red blood cell—is still incomplete, yet we are reasonably sure of where and how the major components are arranged. The so-called integral membrane proteins are the most likely candidates to serve as receptors and transport sites of surface membranes, and these may be organized as macromolecular complexes with parts of their polypeptides in contact with both sides of the bilayer. These complexes may be either mobile or fixed, and if the latter, perhaps by specific interactions with other membrane proteins.

The exposed segments of such glycoproteins could have recognition sites that confer cell, tissue, and/or species specificity created by oligosaccharides attached to the peptide backbone, or by specific amino acid sequences, or a combination of both.

Literature Cited

1. Branton, D. 1969. Membrane structure. *Ann. Rev. Plant Physiol.* 20:209
2. Edidin, M. 1974. Rotational and translational diffusion in membranes. *Ann. Rev. Biophys. Bioeng.* 3:179–201
3. Zwaal, R. F. A., Roelofsen, B., Colley, C. M. 1973. Localization of red cell membrane constituents. *Biochim. Biophys. Acta. 300:159–82*
4. Fairbanks, B., Steck, T. L., Wallach, D. F. H. 1971. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* 10:2606–17
5. Singer, S. J. 1974. The molecular organization of membranes. *Ann. Rev. Biochem.* 43:805–33
6. Marchesi, V. T., Steers, E. 1968. Selective solubilization of a protein component of the red cell membrane. *Science* 159:203–4
7. Juliano, R. L. 1973. The proteins of the erythrocyte membrane. *Biochim. Biophys. Acta.* 300:341–78
8. Steck, T. L. 1974. The organization of proteins in the human red blood cell membrane. *J. Cell Biol.* 62:1–19
9. Marchesi, V. T. 1975. Biochemistry of cell walls and membranes. In *MTP International Review of Science Biochemistry*, ed. C. F. Fox, Ser. 1. Vol. 2., p. 123. Baltimore: Univ. Park
10. Tomita, M., Marchesi, V. T. 1975. Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. *Proc. Natl. Acad. Sci. USA* 72:2964–68
11. Kornfeld, R., Kornfeld, S. 1970. The structure of a phytohemagglutinin receptor site from human erythrocytes. *J. Biol. Chem.* 245:2536–45
12. Thomas, D. B., Winzler, R. J. 1971. Structure of glycoproteins of human erythrocytes. *Biochem. J.* 124:55–59
13. Bretscher, M. S. 1971. Major human erythrocyte glycoprotein spans the cell membrane. *Nature New Biol.* 231:229–32
14. Segrest, J. P., Kahane, I., Jackson, R. L., Marchesi, V. T. 1973. Major glycoprotein of the human erythrocyte membrane evidence for an amphipathic molecular structure. *Arch. Biochem. Biophys.* 155:167–83
15. Morrison, M., Mueller, T. J., Huber, C. T. 1974. Transmembrane orientation of the glycoproteins in normal human erythrocytes. *J. Biol. Chem.* 249:2658–60
16. Shin, B. C., Carraway, K. L. 1974. Lactoperoxidase labeling of erythrocyte membranes from the inside and outside. *Biochim. Biophys. Acta.* 345:141
17. Reichstein, E., Blostein, R. 1975. Arrangement of human erythrocyte membrane proteins. *J. Biol. Chem.* 250:6256–63
18. Cotmore, S., Furthmayr, H., Marchesi, V. T. 1977. Immunochemical evidence for the transmembrane orientation of glycophorin A: localization of ferritin-antibody conjugates in intact cells. *J. Mol. Biol.* 113:539–53
19. Marton, L. S. G., Garvin, J. E. 1973. Subunit structure of the major human erythrocyte glycoprotein: depolymerization by heating ghosts with sodium dodecyl sulfate. *Biochem. Biophys. Res. Commun.* 52:1457–62
20. Furthmayr, H., Marchesi, V. T. 1976. Subunit structure of human erythrocyte glycophorin A. *Biochemistry* 15:1137–44
21. Segrest, J. P., Jackson, R. L., Andrews, E. P., Marchesi, V. T. 1971. Human erythrocyte membrane glycoprotein: a re-evaluation of the molecular weight as determined by SDS polyacrylamide gel electrophoresis. *Biochem. Biophys. Res. Commun.* 44:390–95
22. Silverberg, M., Furthmayr, H., Marchesi, V. T. 1976. The effect of carboxymethylating a single methi-

onine residue on the subunit interaction of glycophorin A. *Biochemistry* 15:1448–54
23. Frye, L. D., Edidin, M. 1970. The rapid intermixing of cell surface antigens after formation of mouse human heterakaryons. *J. Cell Sci.* 7:319–35
24. Tillack, T. W., Scott, R. E., Marchesi, V. T. 1972. The structure of erythrocyte membranes studies by freeze etching. *J. Exp. Med.* 135:1209–27
25. Nicolson, G. L., Singer, S. J. 1974. The distribution and asymmetry of mammalian cell surface saccharides utilizing ferritin-conjugated plant agglutinins as specific saccharide stains. *J. Cell Biol.* 60:236–48
26. Elgsaeter, A., Branton, D. 1974. Intramembrane particle aggregation in erythrocyte ghosts. *J. Cell Biol.* 63:1018–36
27. Nicolson, G. L., Painter, R. G. 1973. Anionic sites of human erythrocyte membranes. *J. Cell Biol.* 59:395–406
28. Elgsaeter, A., Shotton, D. M., Branton, D. 1976. Intramembrane particle aggregation in erythrocyte. *Biochim. Biophys. Acta.* 426:101–22
29. Steck, T. L. 1972. Cross-linking the major proteins of the isolated erythrocyte membrane. *J. Mol. Biol.* 66:295–305
30. Shotton, D. M., Elgsaeter, A., Branton, D. 1975. The influence of spectrin on particle aggregation, surface protein distribution and lipid vesicle blebbing in erythrocyte ghosts. *J. Cell Biol.* 67:397a
31. Yu, J., Branton, D. 1976. Reconstitution of intramembrane particles in recominants of erythrocyte protein band 3 and lipid: effects of spectrin-action association. *Proc. Natl. Acad. Sci. USA* 73:3891–95
32. Bennett, V., Branton, D. 1977. Selective association of spectrin with the cytoplasmic surface of human erythrocyte plasma membranes. *J. Biol. Chem. 252:2753–63*
33. Weed, R. I., Lacelle, P. L. 1969. *Red Cell Membrane: Structure and Function.* ed. G. A. Jamieson, T. J. Greenwalt, pp. 318–338. Philadelphia: Lippincott
34. Weed, R. I., Lacelle, P. L., Merrill, E. W. 1969. Metabolic dependence of red cell deformability. *J. Clin. Invest.* 48:795–809
35. Kirkpatrick, F. H. 1976. Spectrin: Current understanding of its physical, biochemical, and functional properties. *Life Sci.* 19:1–18
36. Sheetz, M. P., Singer, S. J. 1977. On the mechanism of ATP-induced shape changes in human erythrocyte membranes. I. The role of spectrin complex. *J. Cell Biol.* 73:638–46
37. Birchmeier, W., Singer, S. J. 1977. On the mechanism of ATP-induced shape changes in human erythrocyte membranes. II. The role of ATP. *J. Cell Biol.* 73:647–59
38. Gorter, E., Grendel, F. 1925. On bimolecular layers of lipoids on the chromocytes of the blood. *J. Exp. Med.* xli:439–43

CUMULATIVE INDEXES

CONTRIBUTING AUTHORS VOLUMES 25 - 29

A

Aberdeen, E., 26:451-64
Ahmed, S. S., 28:393-409
Ahn, Y. S., 28:299-309
Alexander, P., 27:207-24
Alfrey, A. C., 29:93-98
Annas, G. J., 29:9-14
Anthony, B. F., 28:355-69
Arnold, A. B., 28:187-94
Arnone, A., 25:123-30
Atkinson, A. J., 25:99-113
Atkinson, J. P., 28:291-98

B

Bagshaw, M. A., 26:567-88
Bailey, J. C., 29:417-26
Baldwin, D. S., 27:49-55
Baldwin, R. W., 27:151-63
Barchas, J. D., 26:499-520
Barker, W. H. Jr., 25:75-81
Barnes, A. C., 27:1-7
Becker, M. A., 25:15-28
Begemann, H., 28:131-41
Beisel, W. R., 26:9-20
Berman, M. L., 26:391-401
Bernard, J., 25:39-51
Betts, R. F., 28:103-10
Blum, C. B., 26:365-71
Bonadonna, G., 29:149-75
Bone, H. G. III, 28:111-17
Boshes, B., 26:465-70
Boyar, R. M., 29:509-20
Branch, R. A., 29:205-18
Braverman, L. E., 26:443-49
Brenner, B. M., 28:91-102
Brewer, G. J., 25:29-38
Bristow, J. D., 29:475-83
Brummer, D. L., 25:115-22
Brunner, E. A., 26:391-401
Buckner, C. D., 25:179-98

C

Cadoret, R. J., 26:21-25
Cannon, P. J., 27:485-521
Caro, W. A., 28:251-69
Cavalieri, R. R., 28:57-65
Chase, T. N., 26:419-26
Cheng, S. C., 26:391-401
Cohen, J. J., 29:521-41
Cohen, S., 26:373-90
Collen, D., 25:447-55
Collinsworth, K. A., 25:143-48
Colvin, R. B., 29:191-203
Conklin, R. H., 28:167-77
Craig, W. A., 27:287-300
Crosnier, J., 29:67-72

D

Dale, D. C., 26:277-81
Davison, R., 25:99-113
Dean, R. H., 27:397-405
Deftos, L. J., 25:323-31
DeLuca, H. F., 25:349-67
Dement, W. C., 27:465-84
DeVries, W. C., 26:203-17
Dixon, F. J., 25:83-98
Doherty, J. E., 26:159-71
Downing, D. T., 26:27-32
Dudrick, S. J., 28:517-28
DuPont, H. L., 28:167-77
Durie, B. G. M., 26:283-88
Durrer, D., 27:63-68
Dustan, H. P., 29:485-93

E

Edgington, T. S., 25:303-14
Ehrenkrantz, N. J., 25:9-14
Eickhoff, T. C., 28:411-23
Elharrar, V., 29:417-26
Elin, R. J., 27:127-41
Elliott, G. R., 26:499-520
Epstein, C. J., 29:117-28
Ettinger, P. O., 28:393-409
Ettlinger, R. E., 29:15-22
Evenson, M. A., 28:85-90

F

Fanelli, G. M. Jr., 28:349-54
Fauci, A. S., 26:277-81
Fefer, A., 25:179-98
Feinstone, S. M., 29:359-66
Feldman, D., 26:83-90
Fernstrom, J. D., 25:1-8
Fisher, R. S., 26:373-90
Fitzgerald, F., 29:177-89
Forget, B. G., 26:345-51
Forstot, S. L., 28:21-35
Frank, M. M., 28:291-98
Freeman, H. J., 29:99-116
Freeman, R. G., 27:77-87
Friedman, R. C., 27:57-62
Froelicher, V. F. Jr., 28:1-12

G

Gabbay, K. H., 26:521-36
Gadek, J., 26:291-98
Gangarosa, E. J., 25:75-81
Gasparini, M., 29:149-75
Gault, M. H., 27:117-25
Gavras, H., 27:485-521
Gennari, F. J., 29:521-41
Gerald, P. S., 29:563-70
Gerich, J. E., 27:379-88
Gibbons, R. J., 26:121-36
Giebink, G. S., 29:285-306
Gillis, R. A., 29:73-79
Glatstein, E., 29:81-91
Glenner, G. G., 25:131-35
Gluck, F. W., 27:397-405
Glucksberg, H., 25:179-98
Golbus, M. S., 29:117-28
Goldberg, M., 26:537-50
Goldfine, I. D., 29:407-15
Goldreyer, B. N., 26:219-28
Goldsmith, R. S., 27:181-90
Goodnight, J. E. Jr., 29:231-83
Gottlieb, L. S., 25:411-29
Graber, S. E., 29:51-66
Grams, R. R., 27:199-206
Green, R., 27:57-62
Greenblatt, D. J., 27:407-20
Grossman, M., 28:43-55

Grossman, M. I., 29:495-507
Grumet, F. C., 28:425-52
Guillemin, R., 27:379-88
Guilleminault, C., 27:465-84
Gunnar, R. M., 26:471-83
Gunnells, J. C. Jr., 26:259-75
Gutterman, J. U., 28:489-515 515

H

Habener, J. F., 29:327-42
Hackett, T. P., 27:371-78
Haider, B., 28:393-409
Halmi, K. A., 29:137-48
Hamburger, J., 29:67-72
Hanshaw, J. B., 28:103-10
Harker, L. A., 25:383-400
Harmel, M. H., 26:137-41
Harrington, W. J., 28:299-309
Harrison, D. C., 25:143-48
Hartmann, R. C., 28:187-94
Havel, R. J., 28:195-209
Henderson, E. S., 26:289-306
Henkin, R. E., 26:193-201
Herndon, J. H. Jr., 27:77-87
Hersh, E. M., 28:489-515
Higby, D. J., 26:289-306
Holick, M. F., 25:349-67
Holland, J. F., 28:475-88
Hollister, L. E., 27:431-46
Holman, R. B., 26:499-520
Holmberg, L., 26:33-44
Hook, E. W., 28:371-91
Hopewell, P. C., 27:343-56
Hough, P. V. C., 25:431-46
Hoyumpa, A. M. Jr., 29:205-18
Hunder, G. G., 29:15-22
Hurst, M. W., 27:301-12

I

Ingbar, S. H., 26:443-49

J

Jacquillat, C., 25:39-51
Janowitz, H. D., 27:269-85
Jenkins, C. D., 27:301-12; 29:543-62
Johannes, R. S., 26:589-601
Johnson, M. T., 25:251-67
Jones, S. R., 25:251-67
Jordan, P. H. Jr., 27:191-97

K

Kane, J. J., 26:159-71
Kaplan, B. S., 27:117-25
Kaplan, N. M., 29:31-40
Kattus, A. A., 27:69-75
Kaufman, H. E., 28:21-35
Kelly, J. F., 25:53-68
Kennedy, B. J., 26:235-43
Kim, Y., 29:319-25
Kim, Y. S., 29:99-116
Kinney, D. K., 29:459-73
Klassen, J., 27:117-25
Klein, M. S., 27:89-108
Klein, P. A., 28:311-27
Klerman, G. L., 25:199-217
Koffler, D., 25:149-64
Korenman, S. G., 27:357-70
Krantz, S. B., 29:51-66
Kunin, C. M., 27:287-300

L

Lamberti, J. J., 29:571-82
Langer, G. A., 28:13-20
Lanning, C. F., 26:137-41
Lasseter, K. C., 25:69-74
Le Quesne, L. P., 26:63-74
Lerner, K. G., 25:179-98
Levey, G. S., 25:69-74
Levy, R. I., 26:365-71
Linnemann, C. C. Jr., 28:179-85
Lipinski, J., 26:551-65
Lipscomb, K., 28:271-89
Lofgren, S., 27:407-20
Long, J. M. III, 28:517-28
Lopez, M., 27:453-63
Louria, D. B., 25:219-31
Love, R., 28:131-41
Luby, J. P., 25:251-67
Lyons, M. M., 28:393-409

M

Maas, J. W., 29:219-29
Maddox, D. A., 28:91-102
Mahaffey, J. E., 29:327-42
Majerus, P. W., 29:41-49
Marchesi, V. T., 29:593-603
Marcus, R., 27:357-70
Matthysse, S., 26:551-65; 29:459-73
Mavligit, G. M., 28:489-515
McCabe, W. R., 27:335-41
McCarty, D. J., 25:279-88
McCluskey, R. T., 29:191-203
McConnell, R. B., 25:165-78
McDevitt, H. O., 28:425-52
McGraw, C. P., 26:321-29
McGuffin, W. L. Jr., 26:259-75
McGuigan, J. E., 25:233-49; 29:307-18
McGuire, W. L., 26:353-63
McPhaul, J. J. Jr., 28:37-42
McRee, D. I., 27:109-15
Mello, N. K., 27:321-33
Mendelson, J. H., 27:321-33
Merimee, T. J., 25:137-42
Merin, R. G., 28:75-83
Merrill, J. P., 29:343-58
Messner, R. P., 26:181-92
Michael, A. F., 29:319-25
Miletich, J. P., 29:41-49
Millman, B. S., 28:223-34
Mitchell, J. H., 25:369-81
Modell, J. H., 29:1-8
Moran, J. F., 26:471-83
Morton, D. L., 29:231-83
Moylan, J. A., 28:85-90
Mullins, C. B., 28:271-89
Murray, J. F., 27:343-56
Murray, T., 26:537-50

N

Nasenbeny, J., 28:179-85
Nathan, D. G., 26:345-51
Neer, R., 25:323-31
Neiman, P., 25:179-98
Nicoloff, D. M., 26:143-57
Nilsson, I. M., 26:33-44
Noël, L.-H., 29:67-72
Norman, P. S., 26:337-44

O

Oates, J. A., 29:23-29
Odell, W. D., 29:379-406
Okada, D. M., 28:355-69
Oken, D. E., 26:307-19
Oldewurtel, H. A., 28:393-409
Oliver, J. A., 27:485-521
Ontjes, D. A., 27:263-68
Oratz, M., 26:91-104
Orci, L., 28:119-30
O'Reilly, R. A., 27:245-61

P

Pak, C. Y. C., 28:111-17
Palmer, R. F., 25:69-74
Parisi, A. F., 28:159-66
Patterson, R., 25:53-68

Paulson, G. W., 26:75-81
Perloff, M., 28:475-88
Pierce, A. K., 28:211-21
Plotkin, S. A., 29:583-91
Pochi, P. E., 26:27-32
Pollack, R. E., 25:431-46
Portlock, C. S., 29:81-91
Portnoy, B. L., 28:167-77
Price, M. R., 27:151-63
Prien, E. L. Jr., 26:173-79
Purcell, R. H., 29:359-66

Q

Quest, J. A., 29:73-79
Quie, P. G., 29:285-306
Quinn, J. L. III, 26:193-201

R

Rackley, C. E., 26:105-20
Raisz, L. G., 29:23-29
Rakowski, H., 27:165-80
Ranganathan, N., 27:165-80
Rapoport, B., 28:57-65
Ray, G. R., 26:567-88
Regan, T. J., 28:393-409
Resnekov, L., 29:571-82
Robertson, G. L., 25:315-22
Robertson, J., 28:211-21
Robins, E., 28:67-73
Rose, R. M., 27:301-12
Rosenberg, R. D., 29:367-78
Rosse, W. F., 26:331-36
Rossi, A., 29:149-75
Rothschild, M. A., 26:91-104
Ruderman, N. B., 26:245-58
Russell, R. O. Jr., 26:105-20
Ryan, W. G., 28:143-52

S

Sachar, D. B., 27:269-85
Sachar, E. J., 27:389-96
Salmon, S. E., 26:283-88
Salvaggio, J., 27:453-63
Sanford, J. P., 25:401-10; 26:485-98; 27:421-29
Sasahara, A. A., 28:159-66
Sasazuki, T., 28:425-52
Schacht, R. G., 27:49-55
Schenker, S., 29:205-18
Schiff, G. M., 27:447-52
Schildkraut, J. J., 25:333-48
Schreiber, S. S., 26:91-104
Schrier, S. L., 25:269-78
Scott, H. W. Jr., 27:397-405
Seegmiller, J. E., 25:15-28
Sever, J. L., 26:589-601
Seyberth, H. W., 29:23-29
Shader, R. I., 27:407-20
Shah, P. M., 28:235-50
Shah, S. A., 29:563-70
Sharma, G. V. R. K., 28:159-66
Sherlock, S., 27:143-49
Sherwin, I., 27:37-47
Shoulson, I., 26:419-26
Shull, H. J., 27:397-405
Silen, W., 27:9-22
Silver, M. D., 27:165-80
Sipes, J. N., 28:371-91
Skillman, J. J., 27:9-22
Smith, R. B. III, 26:229-34
Smith, R. T., 28:311-27
Snyder, W. H. III, 28:111-17
Sobel, B. E., 27:89-108
Soll, A. H., 29:495-507
Spitzer, R. L., 27:57-62
Stephen, C. R., 28:153-57
Storb, R., 25:179-98
Stossel, T. P., 29:427-57
Strauss, J. S., 26:27-32
Sweeney, D. R., 29:219-29

T

Tancini, G., 29:149-75
Tarazi, R. C., 29:485-93
Terry, W. D., 25:131-35
Theml, H., 28:131-41
Thomas, E. D., 25:179-98
Thompson, R. L., 28:371-91
Tilkian, A., 27:465-84
Tomasi, T. B. Jr., 28:453-65
Toole, J. F., 26:321-29
Tow, D. E., 28:159-66
Trey, C., 25:411-29
Twito, D. I., 26:235-43

U

Underwood, L. E., 26:427-41
Unger, R. H., 28:119-30
Utiger, R. D., 25:289-302

V

van Houte, J., 26:131-36
Van Wyk, J. J., 26:427-41
Ventura, A. K., 25:9-14
Vermylen, J., 25:447-55
Verstraete, M., 25:447-55
Voorhees, J. J., 28:467-73

W

Wallace, W. A., 26:1-7
Walzer, S., 29:563-70
Ward, L. E., 29:15-22
Warren, W. D., 26:229-34
Waterson, A. P., 27:23-35
Wechsler, A. S., 28:341-48
Weil, M., 25:39-51
Weitzman, E. D., 27:225-43
Wells, S. A. Jr., 27:263-68
Wender, P. H., 26:45-62
Wessler, S., 27:313-19
Wigle, E. D., 27:165-80
Wiktor, T., 29:583-91
Wildenthal, K., 25:369-81
Wilfert, C. M., 29:129-36
Williams, R. C. Jr., 26:181-92
Wilson, C. B., 25:83-98
Wilson, D. R., 28:329-39
Winkelmann, R. K., 28:251-59
Winokur, G., 26:21-25
Wisneski, J. A., 29:475-83
Wolfe, W. G., 26:203-17
Wolff, S. M., 26:277-81; 27:127-41
Wolfsen, A. R., 29:379-406
Woodrow, J. C., 26:165-78

Y

Yen, S. S. C., 26:403-17
Yu, P. N., 26:1-7

Z

Zieve, L., 26:143-57
Zimmerman, T. S., 25:303-14
Zipes, D. P., 29:417-26

CHAPTER TITLES VOLUMES 25-29

ALLERGY AND IMMUNOLOGY		
Animal Models of the Asthmatic State	R. Patterson, J. F. Kelly	25:53-68
Characterization of Amyloid	G. G. Glenner, W. D. Terry	25:131-35
Immunopathogenesis of Systemic Lupus Erythematosus	D. Koffler	25:149-64
Alterations in T- and B-Cells in Human Disease States	R. C. Williams Jr., R. P. Messner	26:181-92
The Antiglobulin Test in Autoimmune Hemolytic Anemia	W. F. Rosse	26:331-36
Immunotherapy (Densensitization) in Allergic Disease	P. S. Norman	26:337-44
IgA-Associated Glomerulonephritis	J. J. McPhaul Jr.	28:37-42
The Current Status of BCG Immunization Against Tuberculosis	T. C. Eickhoff	28:411-23
The Association of Genes in the Major Histocompatibility Complex and Disease Susceptibility	T. Sasazuki, F. C. Grumet, H. O. McDevitt	28:425-52
Structure and Function of Alpha-Fetoprotein	T. B. Tomasi Jr.	28:453-65
BCG as Adjuvant Immunotherapy for Neoplasia	E. M. Hersh, J. U. Gutterman, G. M. Mavligit	28:489-515
Polymyalgia Rheumatica and Giant Cell Arteritis	R. E. Ettlinger, G. G. Hunder, L. E. Ward	29:15-22
ANESTHESIOLOGY		
Ketamine Anesthesia	C. F. Lanning, M. H. Harmel	26:137-41
Effects of Anesthesia on Intermediary Metabolism	E. A. Brunner, S. C. Cheng, M. L. Berman	26:391-401
Effects of Anesthetic Drugs on Myocardial Performance in Man	R. G. Merin	28:75-83
Malignant Hyperpyrexia	C. R. Stephen	28:153-57
Acupuncture: Context and Critique	B. S. Millman	28:223-34
Biology of Drowning	J. H. Modell	29:1-8
CARDIOLOGY		
Sulfonylureas and the Heart	G. S. Levey, K. C. Lasseter, R. F. Palmer	25:69-74
Diphenylhydantoin as an Antiarrhythmic Drug	A. J. Atkinson, R. Davison	25:99-113
Antiarrhythmic Actions of Lidocaine	D. C. Harrison, K. A. Collinsworth	25:143-48
Static (Isometric) Exercise and the Heart: Physiological and Clinical Considerations	J. H. Mitchell, K. Wildenthal	25:369-81
Sudden Death and the Pre-Hospital Phase of Acute Myocardial Infarction	W. A. Wallace, P. N. Yu	26:1-7
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Left Ventricular Function in Acute and Chronic Coronary Artery Disease	C. E. Rackley, R. O. Russell Jr.	26:105-20
Clinical Pharmacology of Digitalis Glycosides	J. E. Doherty, J. J. Kane	26:159-71
Mechanisms of Supraventricular Tachycardias	B. N. Goldreyer	26:219-28
Selective Distal Splenorenal Shunt for Bleeding Esophageal Varices	R. B. Smith III, W. D. Warren	26:229-34

Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259–75
The Steal Syndromes	J. F. Toole, C. P. McGraw	26:321–29
The Total Correction of Congenital Heart Disease	E. Aberdeen	26:451–64
Pathophysiology of Heart Block	J. F. Moran, R. M. Gunnar	26:471–83
Medical and Surgical Treatment of the Wolff-Parkinson-White Syndrome	D. Durrer	27:63–68
Coronary Artery Spasm: A New Appraisal	A. A. Kattus	27:69–75
Medical Management of Myocardial Infarction	M. S. Klein, B. E. Sobel	27:89–108
Mitral Valve Prolapse	E. D. Wigle, H. Rakowski, N. Ranganathan, M. D. Silver	27:165–80
The Detection of Asymptomatic Coronary Artery Disease	V. F. Froelicher Jr.	28:1–12
Ionic Basis of Myocardial Contractility	G. A. Langer	28:13–20
Hypertrophic Cardiomyopathy	P. M. Shah	28:235–50
Critique of Coronary Artery Bypass Surgery	C. B. Mullins, K. Lipscomb	28:271–89
Development of Coronary Collateral Circulation	A. S. Wechsler	28:341–48
Prophylaxis of Infective Endocarditis: A Reevaluation	J. N. Sipes, R. L. Thompson, E. W. Hook	28:371–91
The Role of Ethanol in Cardiac Disease	T. J. Regan, P. O. Ettinger, B. Haider, S. S. Ahmed, H. A. Oldewurtel, M. M. Lyons	28:393–409
Slow-Channel Depolarization: Mechanism and Control of Arrhythmias	J. C. Bailey, V. Elharrar, D. P. Zipes	29:417–26
Left Ventricular Stiffness	J. A. Wisneski, J. D. Bristow	29:475–83
Cardiogenic Hypertension	H. P. Dustan, R. C. Tarazi	29:485–93
Cardiac Assist Devices	J. J. Lamberti, L. Resnekov	29:571–82
CLINICAL PHARMACOLOGY		
Tardive Dyskinesia	G. W. Paulson	26:75–81
Clinical Pharmacology of Digitalis Glycosides	J. E. Doherty, J. J. Kane	26:159–71
Analgesic Abuse and Renal Disease	T. Murray, M. Goldberg	26:537–50
Rational Psychopharmacology for Patients with Medical Diseases	D. J. Greenblatt, R. I. Shader, S. Lofgren	27:407–20
Effective Use of Analgesic Drugs	L. E. Hollister	27:431–46
Neural Actions of Digitalis	R. A. Gillis, J. A. Quest	29:73–79
The Disposition and Effects of Sedatives and Analgesics in Liver Disease	A. M. Hoyumpa Jr., R. A. Branch, S. Schenker	29:205–18
DERMATOLOGY		
Skin Lipids and Acne	J. S. Strauss, P. E. Pochi, D. T. Downing	26:27–32
Human Disease Associated with Exposure to Light	J. H. Herndon Jr., R. G. Freeman	27:77–87
Current Problems in Mycosis Fungoides and Sezary Syndrome	R. K. Winkelmann, W. A. Caro	28:251–69
Pathophysiology of Psoriasis	J. J. Voorhees	28:467–73
ENDOCRINOLOGY AND METABOLISM		
Isolated Growth Hormone Deficiency and Related Disorders	T. J. Merimee	25:137–42
Serum Triiodothyronine in Man	R. D. Utiger	25:289–302
Vasopressin in Osmotic Regulation in Man	G. L. Robertson	25:315–22
Vitamin D Metabolism	M. F. Holick, H. F. DeLuca	25:349–67
Metabolic Response to Infection	W. R. Beisel	26:9–20
The Role of Hormone Receptors in the Action of Adrenal Steroids	D. Feldman	26:83–90

Regulation of Albumin Metabolism	M. A. Rothschild, M. Oratz, S. S. Schreiber	26:91-103
Muscle Amino Acid Metabolism and Gluconeogenesis	N. B. Ruderman	26:245-58
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
Endocrine Therapy of Breast Cancer	W. L. McGuire	26:353-63
Interconversions of Apolipoprotein Fragments	C. B. Blum, R. I. Levy	26:365-71
Effects of Anesthesia on Intermediary Metabolism	E. A. Brunner, S. C. Cheng, M. L. Berman	26:391-401
Gonadotropin-Releasing Hormone	S. S. C. Yen	26:403-17
Relation Between Growth Hormone and Somatomedin	J. J. Van Wyk, L. E. Underwood	26:427-41
Active Form of the Thyroid Hormone and Somatomedin	S. H. Ingbar, L. E. Braverman	26:443-49
Hyperglycemia, Polyol Metabolism, and the Complications of Diabetes Mellitus	K. H. Gabbay	26:521-36
Potential Microwave Injuries in Clinical Medicine	D. I. McRee	27:109-15
The Effects of Calcium and Phosphorus in Hemodialysis	R. S. Goldsmith	27:181-90
Circadian Rhythms and Episodic Hormone Secretion in Man	E. D. Weitzman	27:225-43
Estrogens and the Human Male	R. Marcus, S. G. Korenman	27:357-70
Somatostatin: Physiological and Clinical Significance	R. Guillemin, J. E. Gerich	27:379-88
Medical Aspects of Riot Control (Harassing) Agents	J. P. Sanford	27:421-29
Impaired Peripheral Conversion of Thyroxine to Triiodthyronine	R. R. Cavalieri, B. Rapoport	28:57-65
Diagnosis of Hyperparathyroidism	H. C. Bone III, W. H. Snyder, C. Y. C. Pak	28:111-17
The Role of Glucagon in the Endogenous Hyperglycemia of Diabetes Mellitus	R. H. Unger, L. Orci	28:119-30
Paget's Disease of Bone	W. G. Ryan	28:143-52
Classification of the Hyperlipidemias	R. J. Havel	28:195-209
Prostaglandins and Hypercalcemic States	H. W. Seyberth, L. G. Raisz, J. A. Oates	29:23-29
Clinical Hypophosphatemia	F. Fitzgerald	29:177-89
Osteomalacia and Disorders of Vitamin D Metabolism	J. F. Habener, J. E. Mahaffey	29:327-42
Humoral Syndromes Associated with Cancer	W. D. Odell, A. R. Wolfsen	29:379-406
Medical Treatment of Acromegaly	I. D. Goldfine	29:407-15
Control of the Onset of Puberty	R. M. Boyar	29:509-20
GASTROINESTINAL TRACT		
Role of Gastrin in Hypersecretory Disorders in Man	J. E. McGuigan	25:233-49
Regulation in Albumin Metabolism	M. A. Rothschild, M. Oratz, S. S. Schreiber	26:91-104
Pathogenesis of Hepatic Coma	L. Zieve, D. M. Nicoloff	26:143-57
Selective Distal Splenorenal Shunt for Bleeding Esophageal Varices	R. B. Smith III, W. D. Warren	26:229-34
Disorders of the Lower Esophageal Sphincter	R. S. Fisher, S. Cohen	26:373-90
Stress Ulceration in the Acutely Ill	J. J. Skillman, W. Silen	27:9-22
Infectious Particles in Hepatitis	A. P. Waterson	27:23-35
Parietal Cell Vagotomy Without Drainage for the Treatment of Duodenal Ulcer	P. H. Jordan Jr.	27:191-97
New Observations in Crohn's Disease	H. D. Janowitz, D. B. Sachar	27:269-85
Metabolic Complications of Jejunoileal Bypass Operations for Morbid Obesity	H. W. Scott Jr., R. H. Dean,	

	H. J. Shull, F. W. Gluck	27:397-405
Viral Agents and Diarrheal Illness	H. L. DuPont, L. Portnoy, R. H. Conklin	28:167-77
Digestion and Absorption of Protein	H. J. Freeman, Y. S. Kim	29:99-116
Gastrointestinal Hormones	J. E. McGuigan	29:307-18
Non-A, Non-B Hepatitis	S. M. Feinstone, R. H. Purcell	29:359-66
Cellular Mechanisms in Acid Secretion	A. H. Soll, M. I. Grossman	29:495-507
HEMATOLOGY		
2,3-DPG and Erythrocyte Oxygen Affinity	G. J. Brewer	25:29-38
Treatment of Acute Granulocytic Leukemias	J. Bernard, M. Weil, C. Jacquillat	25:39-51
Mechanism of Action of Hemoglobin	A. Arnone	25:123-30
Immuno-Prevention of Rh Hemolytic Disease of the Newborn	R. B. McConnell, J. C. Woodrow	25:165-78
Marrow Transplantation for Aplastic Anemia and Acute Leukemia	P. Neiman, E. D. Thomas, C. D. Buckner, R. Storb, A. Fefer, H. Glucksberg, K. G. Lerner	25:179-98
Clinical Utility of In Vitro Evaluation of Granulopoiesis	P. L. Greenberg, S. L. Schrier	25:269-78
Molecular Immunology of Factor VIII	T. S. Zimmerman, T. S. Edgington	25:303-14
Control of Platelet Production	L. A. Harker	25:383-400
Intravascular Coagulation in Liver Disease	M. Verstraete, J. Vermylen, D. Collen	25:447-55
Von Willebrand's Disease	L. Holmberg, I. M. Nilsson	26:33-44
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Cellular Kinetics, Staging, and Immunoglobulin Synthesis in Multiple Myeloma	B. G. M. Durie, S. E. Salmon	26:283-88
Granulocyte Transfusion Therapy	D. J. Higby, E. S. Henderson	26:289-306
The Antiglobulin Test in Autoimmune Hemolytic Anemia	W. F. Rosse	26:331-36
Thalassemia	B. G. Forget, D. G. Nathan	26:345-51
Hemochromatosis: Course and Treatment	S. Sherlock	27:143-49
Vitamin K and the Oral Anticoagulant Drugs	R. A. O'Reilly	27:245-61
Medical Management of Venous Thrombosis	S. Wessler	27:313-19
Factors in the Pathomechanism of Chronic Lymphocytic Leukemia	H. Theml, R. Love, H. Begemann	28:131-41
Paroxysmal Nocturnal Hemoglobinuria (PNH) as a Clonal Disorder	R. C. Hartmann, A. B. Arnold	28:187-94
Cold Agglutinins and Cold-Agglutinin Disease	M. M. Frank, J. P. Atkinson, J. Gadek	28:291-98
Treatment of Idiopathic Thrombocytopenic Purpura (ITP)	Y. S. Ahn, W. J. Harrington	28:299-309
Relationships Between Platelets and Coagulation Factors in Hemostasis	P. W. Majerus, J. P. Miletich	29:41-49
Erythropoietin and the Control of Red Cell Production	S. E. Graber, S. B. Krantz	29:51-66
Heparin, Antithrombin, and Abnormal Clotting	R. D. Rosenberg	29:367-78
Contractile Proteins in Cell Structure and Function	T. P. Stossel	29:427-57
INFECTION		
Venezuelan Equine Encephalitis Virus Infec- in Man	N. J. Ehrenkrantz, A. K. Ventura	25:9-14
Food Poisoning Due to Vibrio Parahaemolyticus	W. H. Barker Jr., E. J. Gangarosa	25:75-81
Preventive Therapy of Tuberculosis	D. L. Brummer	25:115-22
Infectious Complications of Nonalcoholic Drug Abuse	D. B. Louria	25:219-31
Antiviral Chemotherapy	J. P. Luby, M. T. Johnson, S. R. Jones	25:251-67

Metabolic Response to Infection	W. R. Beisel	26:9-20
Granulocyte Transfusion Therapy	D. J. Higby, E. S. Henderson	26:289-306
Urinary Tract Symptoms and Infections	J. P. Sanford	26:485-98
Biology of Endotoxin	R. J. Elin, S. M. Wolff	27:127-40
Significance of Serum Protein and Tissue Binding of Antimicrobial Agents	W. A. Craig, C. M. Kunin	27:287-300
Immunoprophylaxis of Gram-Negative Bacillary Infections	W. R. McCabe	27:335-41
Reye's Syndrome	G. M. Schiff	27:447-52
Cytomegalovirus (CMV) in the Compromised Host(s)	R. F. Betts, J. B. Hanshaw	28:103-10
Pertussis in the Adult	C. C. Linnemann Jr., J. Nasenbeny	28:179-85
The Emergence of Group B Streptococci in Infections of the Newborn Infant	B. F. Anthony, D. M. Okada	28:355-69
E. Coli Meningitis: K_1 Antigen and Virulence	C. M. Wilfert	29:129-36
Otitis Media: The Spectrum of Middle Ear Inflammation	G. S. Geibink, P. G. Quie	29:285-306
Chronic Bacteremia and Nephritis	Y. Kim, A. F. Michael	29:319-25
Rabies Vaccination	S. A. Plotkin, T. Wiktor	29:583-91

KIDNEYS AND URINARY TRACT

Immunopathology and Glomerulonephritis	C. B. Wilson, F. J. Dixon	25:83-98
Calcium Oxalate Renal Stones	E. L. Prien Jr.	26:173-79
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
Acute Renal Failure (Vasomotor Nephropathy): Micropuncture Studies of the Pathogenetic Mechanism	D. E. Oken	26:307-19
Urinary Tract Symptoms and Infections	J. P. Sanford	26:485-98
Analgesic Abuse and Renal Disease	T. Murray, M. Goldberg	26:537-50
Late Sequelae of Poststreptococcal Glomerulonephritis	D. S. Baldwin, R. G. Schacht	27:49-55
Glomerular Injury in Patients with Neoplasia	B. S. Kaplan, J. Klassen, M. H. Gault	27:117-25
Interrelations of Renin, Angiotensin II, and Sodium in Hypertension and Renal Failure	H. Gavras, J. A. Oliver, P. J. Cannon	27:485-521
Glomerular Filtration of Fluid and Macromolecules: The Renal Response to Injury	D. A. Maddox, B. M. Brenner	28:91-102
Renal Function During and Following Obstruction	D. R. Wilson	28:329-39
Urate Excretion	G. M. Fanelli Jr.	28:349-54
Recurrent Glomerulonephritis After Renal Transplantation	J. Hamburger, J. Crosnier, L.-H. Noël	29:67-72
Dialysis Encephalopathy Syndrome	A. C. Alfrey	29:93-98
Immunological Aspects of Renal Tubular and Interstitial Diseases	R. T. McCluskey, R. B. Colvin	29:191-203
Dialysis Versus Transplantation in the Treatment of End-Stage Renal Disease	J. P. Merrill	29:343-58
Renal Tubular Acidosis	F. J. Gennari, J. J. Cohen	29:521-41

MISCELLANEOUS

Genetic Aspects of Gout	M. A. Becker, J. E. Seegmiller	25:15-28
Medical Aspects of Recreation Skin and Scuba Diving	J. P. Sanford	25:401-10
Dental Caries	R. J. Gibbons, J. van Houte	26:121-36
Unusual Etiologies of Fever and Their Evaluation	S. M. Wolff, A. S. Fauci, D. C. Dale	26:277-81
Implications of Mass Automated Instruments on Medical Practice	R. R. Grams	27:199-206

Confidentiality in Medical Practice	M. Grossman	28:43-55
Informed Consent	G. J. Annas	29:9-14
Recent Membrane Research and Its Implications for Clinical Medicine	V. T. Marchesi	29:593-603
MUSCULO-SKELETAL DISORDERS		
Crystal Deposition Joint Disease	D. J. McCarty	25:279-88
NEOPLASIA		
Scanning Techniques to Assess Thyroid Nodules	J. L. Quinn III, R. E. Henkin	26:193-201
Treatment of Testicular Cancer	D. I. Twito, B. J. Kennedy	26:235-43
Cellular Kinetics, Staging, and Immunoglobulin Synthesis in Multiple Myeloma	B. G. M. Durie, S. E. Salmon	26:283-88
Endocrine Therapy of Breast Cancer	W. L. McGuire	26:353-63
The Role of Radiation Therapy in the Definitive Treatment of Adenocarinoma of the Prostate	G. R. Ray, M. A. Bagshaw	26:567-88
Tumor Antigens and Tumor-Host Relationships	R. W. Baldwin, M. R. Price	27:151-63
The Functions of the Macrophage in Malignant Disease	P. Alexander	27:207-24
Multiple Endocrine Neoplasia Type II	S. A. Wells Jr., D. A. Ontjes	27:263-68
The Role of Oncogenic Viruses in Neoplasia	P. A. Klein, R. T. Smith	28:311-27
Surgical Adjuvant Chemotherapy	M. Perloff, J. F. Holland	28:475-88
The Non-Hodgkin's Lymphomas: Current Concepts and Management	C. S. Porlock, E. Glatstein	29:81-91
Chemotherapy in Prevention of the Recurrence of Resectable Cancer	G. Bonadonna, G. Tancini, A. Rossi, M. Gasparini	29:149-75
Immunotherapy for Malignant Disease	J. E. Goodnight Jr., D. L. Morton	29:231-83
NEUROLOGY		
The Minimal Brain Dysfunction Syndrome	P. H. Wender	26:45-62
Tardive Dyskinesia	G. W. Paulson	26:75-81
Pathogenesis of Hepatic Coma	L. Zieve, D. M. Nicoloff	26:143-57
The Steal Syndromes	J. F. Toole, C. P. McGraw	26:321-29
Huntington's Disease	I. Shoulson, T. N. Chase	26:419-26
A Definition of Cerebral Death	B. Boshes	26:465-70
Neuroregulators and Sleep Mechanisms	R. B. Holman, G. R. Elliott, J. D. Barchas	26:499-520
Subacute Sclerosing Panencephalitis	R. S. Johannes, J. L. Sever	26:589-601
Temporal Lobe Epilepsy: Neurological and Behavioral Aspects	I. Sherwin	27:37-47
NUTRITION		
Modification of Brain Serotonin by the Diet	J. D. Fernstrom	25:1-8
Applications and Hazards of Total Intravenous Hyperalimentation	S. J. Dudrick, J. M. Long III	28:517-28
ONCOLOGY		
Medical Management of the Hypercalcemia of Malignancy	L. J. Deftos, R. Neer	25:323-31
The Cell Surface and Malignant Transformation	R. E. Pollack, P. V. C. Hough	25:431-46
OPHTHALMOLOGY AND OTORHINOLARYNGOLOGY		
Corneal Transplantation	S. L. Forstot, H. E. Kaufman	28:21-35
PSYCHIATRY		
Pharmacotherapy of Schizophrenia	G. L. Klerman	25:199-217
Biogenic Amines and Affective Disorders	J. J. Schildkraut	25:333-48
X-Linkage in Manic-Depressive Illness	R. J. Cadoret, G. Winokur	26:21-25
Biochemical Aspects of Schizophrenia	S. Matthysse, J. Lipinski	26:551-65
Reassement of Homosexuality and Transsexualism	R. C. Friedman, R. Green, R. L. Spitzer	27:57-62

The Relation of Psychological Stress to Onset of Medical Illness	M. W. Hurst, C. D. Jenkins, R. M. Rose	27:301-12
Behavioral and Biochemical Interrelations in Alcoholism	J. H. Mendelson, N. K. Mello	27:321-33
Psychological Assistance for the Dying Patient and His Family	T. P. Hackett	27:371-78
Neuroendocrine Dysfunction in Depressive Illness	E. J. Sachar	27:389-96
The Sleep Apena Syndromes	C. Guilleminault, A. Tilkian, W. C. Dement	27:465-84
New Concepts in the Diagnosis of Psychiatric Disorders	E. Robins	28:67-73
Anorexia Nervosa: Recent Investigations	K. A. Halmi	29:137-48
Specificity of Depressive Diseases	D. R. Sweeney, J. W. Maas	29:219-29
Genetic Transmission of Schizophrenia	D. K. Kinney, S. Matthysse	29:459-73
Behavorial Risk Factors in Coronary Artery Disease	C. D. Jenkins	29:543-62
The XYY Genotype	S. Walzer, P. S. Gerald, S. A. Shah	29:563-70
RADIOLOGY		
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Scanning Techniques to Assess Thyroid Nodules	J. L. Quinn III, R. E. Henkin	26:193-201
The Role of Radiation Therapy in the Definitive Treatment of Adenocarcinoma of the Prostate	G. R. Ray, M. A. Bagshaw	26:567-88
REPRODUCTION		
Gonadotropin-Releasing Hormone	S. S. C. Yen	26:403-17
Some Implications of Population Control	A. C. Barnes	27:1-7
Cardiovascular Complications of Oral Contraceptives	N. M. Kaplan	29:31-40
The Prenatal Diagnosis of Genetic Disorders	C. J. Epstein, M. S. Golbus	29:117-28
RESPIRATORY SYSTEM		
Oxygen Toxicity	W. G. Wolfe, W. C. DeVries	26:203-17
The Adult Respiratory Distress Syndrome	P. C. Hopewell, J. F. Murray	27:343-56
Hypersensitivity Pneumonitis: Current Concepts of Etiology and Pathogenesis	M. Lopez, J. Salvaggio	27:453-63
Diagnosis and Treatment of Fat Embolism	J. A. Moylan, M. A. Evenson	28:85-90
Diagnosis of Pulmonary Embolism	G. V. R. K. Sharma, D. E. Tow, A. F. Parisi, A. A. Sasahara	28:159-66
Pulmonary Complications of General Surgery	A. K. Pierce, J. Robertson	28:211-21
THERAPEUTICS AND TOXICOLOGY		
The Effects of Fluorinated Anesthetics on the Liver and Kidneys	L. S. Gottlieb, C. Trey	25:411-29